Investigation on Applying Modular Ontology to Statistical Language Model for Information Retrieval

by

Jiakang Chang

A thesis submitted in partial fulfilment for the requirements of the degree of

Doctor of Philosophy

University of Central Lancashire

February 2015

DECLARATION

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution. I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work.

Signature ____________________________

School: Computing, Engineering and Physical Sciences

TO MY DEAREST FAMILY

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Dr. Zimin Wu and Dr. Peter Gray for their unfailing help, patience and encouragement throughout every stage of this research. The work presented in this thesis could not have been done without their inspirational ideas and valuable insights.

I would like to thank Dr. David George for his invaluable help and suggestions during my PhD study. His PhD work is the solid foundation of this project.

I should most like to thank my family in China, for their support and trust. Without their unfailing love and encouragement, it would have proved much more difficult to get through my research.

This work is dedicated to Almighty God. May His name be glorified now and forever!

Investigation on Applying Modular Ontology to Statistical Language Model for Information Retrieval

ABSTRACT

The objective of this research is to provide a novel approach to improving retrieval performance by exploiting Ontology with the statistical language model (SLM). The proposed methods consist of two major processes, namely ontology-based query expansion (OQE) and ontology-based document classification (ODC). Research experiments have required development of an independent search tool that can combine the OQE and ODC in a traditional SLM-based information retrieval (IR) process using a Web document collection.

This research considers the ongoing challenges of modular ontology enhanced SLM-based search and addresses three contribution aspects. The first concerns how to apply modular ontology to query expansion, in a bespoke language model search tool (LMST). The second considers how to incorporate OQE with the language model to improve the search performance. The third examines how to manipulate such semantic-based document classification to improve the smoothing accuracy. The role of ontology in the research is to provide formally described domains of interest that serve as context, to enhance system query effectiveness.

A Java-based language model search tool (LMST) has been developed to interrogate a large, independent TREC Web document corpus, i.e. WT2g. The experiments will measure the success of ontology enhanced search models by comparing precision outcomes in the 10% to 30% recall range. Performance evaluation will be primarily based on an average of the precision values (APV) in first 30% recall points.

Search experiment outcomes have justified the approached adopted. Empirical results show that established models achieved remarkable improvement over the basic language model. Overall APV results illustrate that an ontology enhanced language model provides improved APV in 96% of 50 query term sets, whilst 4% of query sets were tied results, compared to the baseline. The results also demonstrate 30% of query sets achieved improved recall.

Keywords:

Information Retrieval, Statistical Language Model, Smoothing, Ontology-based Query Expansion, Ontology-based Document Classification.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The objective of this PhD is to provide a novel approach to improve retrieval performance by applying Ontology to statistical language model (SLM). The proposed methods consist of two major processes, namely ontology-based query expansion (OQE) and ontology-based document classification (ODC). Research experiments have required an independent search tool that can combine the OQE and ODC in a traditional SLM based information retrieval (IR) process using a Web document collection.

As online information grows dramatically, robust search engines are playing a signification role in daily life. The determinant for all search engines is the problem of designing an effective retrieval model that can retrieve documents accurately, according to the user's queries. SLM has been widely applied in the IR community in the past decades because of its solid statistical foundation and empirical effectiveness. The principle of SLM based IR is to estimate a language model for each document based on its term probability distribution, and rank the document by the likelihood between query and estimated language model.

According to the statistical principle, the accuracy of a statistical estimator relies on the sampling size of data. It is a challenge to estimate an accurate document language model because maximum likelihood estimation is typically used to produce language model for each document based on a limited amount of the document text. Therefore, the accuracy weakness is compensated for in this work by using smoothing to adjust estimated results. The problem of SLM based IR is essentially reduced to how to "smooth" the document language model.

Traditional smoothing approaches have utilised the term probability distribution in the entire collection with simple interpolation (Chen and Goodman, 1996; Zhai and Lafferty, 2004; Zhai, 2009) and several recent studies demonstrated that document corpus structures can be exploited to improve the search effectiveness (Zhai and Lafferty, 2001; Kurland and Lee, 2004; Liu and Croft, 2004; Mei et al., 2008; Zhou, 2008a). However, these approaches deliver only limited precision in identifying relevant documents and fail to smooth potential relevant documents that contain related terms, but none of the query terms. For example, because the same word may have more than one meaning (polysemy), basic SLM just matches the keyword and irrelevant documents can be retrieved. Conversely, given possible diversity of relevant words, an author may use different words (synonymy) to describe the same topic, thus the relevant document will be missed.

To solve this problem, modular ontology is a feasible way to establish a machine-interpretable relationship between query terms and document. In computing, modular ontology is a formalised vocabulary, of concepts and their relationships, and an explicit assumption of consensual domain knowledge which can serve as a reference point for a related information source. Therefore, modular ontology for information retrieval can extract information more intelligently by semantic association, rather than by simply matching the keyword. Ontology can provide beneficial in a pure statistical approach, where an IR system does no estimate a user's information requirement.

This research considers the ongoing challenges of modular ontology enhanced SLM-based search and addresses three contribution aspects. The first concerns how to apply modular ontology to query expansion, in a bespoke language model search tool (LMST). The second consider how to incorporate OQE with the language model to improve the search performance. The third examines how to manipulate such semantic-based query expansion to document classification to improve the smoothing accuracy. The primary objective is attempt to applying OQE and ODC to existing language model to improve search effectiveness.

A series of search experiments will identify the issue of the keyword query expansion, by ontology traversal, and whether or not OQE and ODC enhanced SLM can exploit the benefits of ontology semantics to improve the search effectiveness. The experiment will assess the success of OQE and ODC against baseline keyword-only with the entire collection (KO+entire) smoothing respectively, by comparing precision outcomes primarily in the early recall points. Moreover, since OQE and ODC can enhance different aspects of the search process, they can be combined to further improve the performance. To provide a consistent approach, performance evaluation will be primarily based on an average of precision values (APV) for the 10%, 20% and 30% recall points.

The research will demonstrate that 53 derived modular ontology enhanced SLMs improve search effectiveness significantly, compared to traditional non-semantic search. Experiment results will show that OQE and ODC combination search model have achieved 23.58% APV improvement (Fig 71).

The rest of this thesis is organised as follows. Chapter 2 will examine current development and search activities in language model, ontology and query expansion in information retrieval. Chapter 3 will propose methodology applied in the experiments. Chapter 4 will introduce the modular ontology enhanced language model in two different ways, i.e. OQE and ODC. Chapter 5 will illustrate the experimentation search process and implementation. Chapter 6 will present and analyse the experiment results and Chapter 7 will summarise and evaluate the outcomes. Finally, we present conclusions and discuss future work in Chapter 8.

2. LITERATURE REVIEW

This chapter attempts to synthesise the issues that characterise the problems in relation to information retrieval, statistical language model, ontology and query expansion. Related work will be considered, in terms of the significance and relevance of the work. The literature review provides the basis for guiding the discussion and justifying the selected search problem: how a semantic enhanced statistical language model might improve retrieval precision and recall, by using ontology-based query expansion and ontology-based document classification.

2.1 Information retrieval

With the inexorable growth of online information, robust search engines are playing a more and more significant role to help users manage and make use of information in different kind of data, such as text documents, pictures and videos. Information retrieval (IR) is the underlying science of search engines, to help a user to find the relevant information from the data collections, to satisfy different information requirements.

2.1.1 Overview

In the 1950s, the initial search engines of information retrieval were designed for library management systems (Spärck Jones, 1997). Over the decades, IR technologies have matured and have been widely used in commercial search engines, i.e. Google, Yahoo and Bing. Existing search engine queries often retrieve a long list of documents, many of which do not always satisfy a user's information requirements. To improve search engine accuracy, one of the most fundamental challenges is to establish a robust retrieval model that can search information effectively.

IR focus on the information level, which is typically presented both diversified and imprecise, as a result Rijsbergen (Rijsbergen, 1979) proposed that the main difference in data retrieval and information retrieval is that the former are usually looking for exact match (relevant or irrelevant) and IR seeks a best match (highly relevant or less relevant).

Fig 1 shows a typical IR procedure proposed by Croft in 1993 (Croft,1993). Squared boxes represent original data and rounded boxes denote basic processes in IR system.

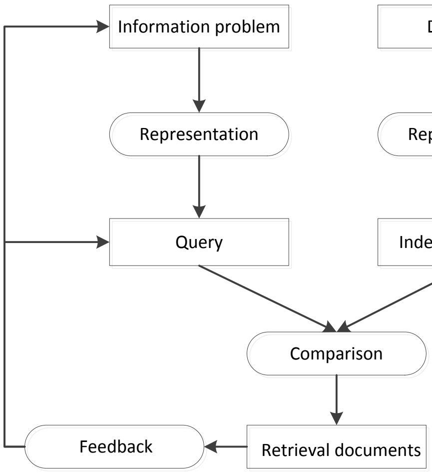


Fig. 1 Basic processing of IR(Croft, 1993)

The procedure contains three main processes, i.e. representation of user's information problem (such as keywords, images and voices), representation of a document to index and comparisons between two components (Hiemstra, 2001).

1. Representing the user's information problem is usually called the query formulation process. The formulation process results in a formal representation of the information problems, i.e. a user's information requirement is reflected in a succinct query.

2. The process of representing the document is referred to the indexing process. Indexing relies on the relevance notion of its underlying retrieval model where documents are represented in different ways, e.g. vectors, probability distribution or logic sets.

3. The comparison process is also called matching processing. The comparison process results in a ranked relevance documents list. The similarity between the

document and the query is determined by the IR search model. It defines the notion of relevance and provides scoring function for document scoring calculation.

The key objective of an IR system is to rank documents to satisfy the user's information needs so that that highly relevant documents are ranked above the less relevant and irrelevant ones (Prabhakar Raghavan 2008; Zhai, 2009). Therefore, the system should able to assign a higher score to the more relevant document than an irrelevant document. Evidently, the effectiveness and accuracy of an IR system is primarily determined by a robust scoring method. The score of the documents relies on a retrieval function, which typically produced based on a retrieval model that defines the notion of relevance and implemented by the system.

2.1.2 Search models

IR search models employ different mathematical mechanisms, to determine the relevance between the document and the query, which include algebra, probability and statistics. Three prominent models are the vector space model (Salton et al., 1975), probabilistic model (Spärck Jones, 1997) and statistical language model (Ponte and Croft, 1998).

In vector space model, documents and query are represented by a vector in the same term vector space respectively, and the vectors are weighted according to adopted weighting approach, e.g. tf-idf. The similarity between the document and the query is produced by cosine similarity between two vectors (Prabhakar Raghavan 2008).

The probabilistic model is focused on the question "What is the probability that the document is relevant to this query?" (Spärck Jones et al., 2000). Given a query, a document is assumed to be either relevant or irrelevant (Robertson, 1977). Therefore, as the search system cannot determine true relevant documents, the document is ranked according to the relevance probability.

In 1988, Ponte and Croft (Ponte and Croft, 1998) first proposed using the statistical language model for document searching. The scoring of each document is determined by the likelihood between the language model of each document and the query, so called query likelihood scoring.

However, because of the limited information observed from the succinct query, the above search models cannot retrieve the document beyond the user's input, i.e. if the users describe their information requirement using *homonymy*, a system is unlikely to return a page containing none of the initial query terms.

On the other hand, the above model is based on a simple assumption, that if a document contains the keywords, it should be treated as the potential relevant one. However, the system cannot filter the document content's *polysemy* noise from an ambiguous text expression.

Therefore, rather than relying solely on existing data presentation approaches, could semantics be used to provide more information about user's requirements, to improve the effectiveness of IR system?

2.1.3 Search effectiveness measure: precision and recall

Precision and recall measurement is typically used to evaluate the search effectiveness of IR model (Rijsbergen, 1979). Precision is the percentage of relevant documents in all retrieved results, and recall is the percentage of all relevant documents that are successfully retrieved.

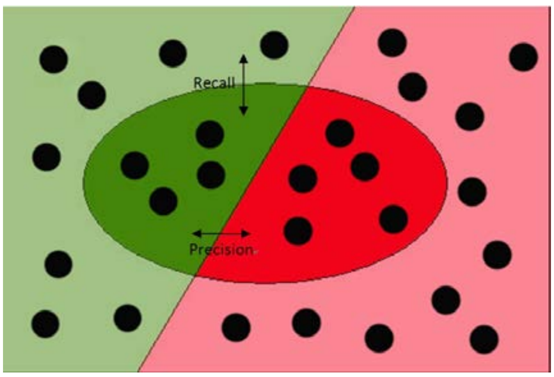


Fig. 2 Precision and Recall (Nichtich, 2008)

Fig 2 represents the definition of precision and recall. The spots indicate the documents in the collection: the green part on the left denotes potential relevant documents in the collection; the red part on the right denotes irrelevant documents and the spots in the centre ellipse represent the retrieved documents. Precision and Recall (P&R) are defined in the following formulas.

$$\text{Precision} = \frac{|\, relevant\ documents\ \cap documents\ retrieved\,|}{|document\ retrieved|}$$
$$\text{Recall} = \frac{|\, relevant\ documents\ \cap documents\ retrieved\,|}{|relevant\ documents|} \tag{1}$$

A determination of search effectiveness in identifying relevant documents can be achieved by applying IR relevance scoring algorithm results in a graph of precision against recall. Because the system will return thousands of the documents in order, typically the retrieval results are evaluated at a given cut-off intervals (Zhai, 2009), i.e. measure the cumulative returned document precision values say for every 10% interval of recall.

2.2 Statistical language model

The accuracy and effectiveness of the retrieval system are directly determined by the quality of the scoring function of IR model adopted. This chapter will review the correlated work of statistical language model (SLM).

2.2.1 Overview

Pioneering work can be traced to 1961. Shannon proposed a mechanism that uses probability to process natural language problems, i.e. a model of data communication system. The fundamental procedures of communication system are shown in Fig 3.

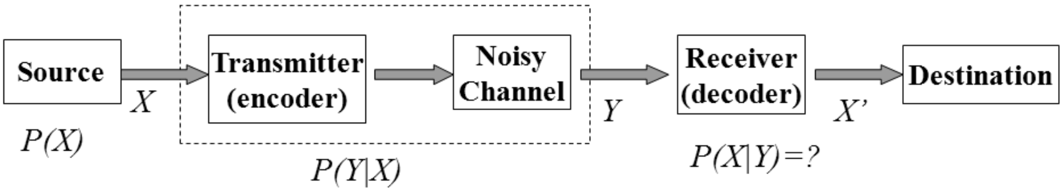


Fig. 3 Fundamental model of communication

Where the *X* is the input of the transmit channel, which is represented in the dashed rectangle, and the *Y* denotes its corresponding output. The task of the communication model is to maximally estimate the initial input data *X* according to the output *Y*. Based on Bayes' theorem, *P (X|Y)*P (Y) = P (Y|X)*P (X)*, since output *Y* is determinate, *P (Y)* is equal to a constant, the equation can be decomposed into:

$$\hat{X} = argmax_x P(X|Y) = argmax_x P(Y|X)P(X) \quad (2)$$

Where, the $\hat{X}$ indicates the information destination which can maximize the probability *P(X|Y)*.

This mechanism was applied in various areas, e.g. the task of speech recognition is to estimate the speech word sequence *X* according to the speech signal *Y*; in machine translation system, *X* represents the initial sentence in one language, *Y* is the result of translation in another language which closest to the content in the original language. Before the SLM been applied to text retrieval, it had already been used successfully in related areas, e.g. synchronous translation and speech translation (Berger and Lafferty, 1999). SLM has also achieved significant improvements in the natural language processing area, e.g. natural language generation and language summarization (Marcu, 1997).

The principle of SLM: *"Statistical language model is a probability distribution P(s) over strings S that attempts to reflect how frequently a string S occurs as a sentence."* is from an article written by Ponte and Croft (Ponte and Croft, 1998), they first advanced the SLM approach to ad hoc textual information retrieval.

The task of SLM is to estimating probability distribution regularities to capture relevant information (McCallum and Nigam, 1998). Specifically in textual information retrieval, it involves estimating a language model for each document based on words' probability distribution and ranks the document reference to its likelihood of query (Ponte and Croft, 1998; Manning et al., 2008; Prabhakar Raghavan 2008; Hiemstra, 2009). For a query *Q* and a document *D*, the communication model formula (2) is refined as:

$$\hat{X} = argmax_D P(D|Q) = argmax_x P(Q|D)P(D) \qquad (3)$$

The prior probability *P (D)* indicates the probability to capture the document from the document corpus and it usually assumed to be a constant, i.e. each document is equally likely to be retrieved. This assumption has been adopted in more recent research (Ponte and Croft, 1998; Baeza-Yates and Ribeiro-Neto, 1999; Zhai and Lafferty, 2004). Instead of assuming equal prior probability of the output, the alternative *P(D)* has been applied to capture non-textual information, e.g. audio-visual data or other components of documents (Corsaro, 1982; Maguire et al., 1999).

A unigram language model is utilised to simplify the calculation of the posterior probability *P (Q|D)*. It assumes that a word sequence/query input is generated by each word independently; the orders of words in a query or documents are not been considered, i.e. bag of words.

Because the sophisticated language model required more data to estimate the parameters, it makes the estimated complex language model inaccurate. So far, the unigram langue model has demonstrated effectiveness and accuracy for IR, and the more sophisticated language model tend not to improve results much over the unigram SLM (Ponte and Croft, 1998; Maguire et al., 1999; Croft, 2003). Therefore, the score of the *D* with regard to *Q* is defined as product value of probability of each keyword in the query.

$$Score\ (D,Q) = P(Q|D) = \prod_{i=1}^{n} P(q_i|D) \qquad (4)$$

Where, q_i is the independent word in the query. The probability of each query term is equal to its relative frequency in D, i.e. maximum likelihood estimation.

$$P(q_i|D) = \frac{c(q_i, D)}{|D|} \tag{5}$$

Where $c(q_i,D)$ is the count of word q_i in D, and $|D|$ is the document length.

In the last decade, the original unigram language model has been extended in several directions. In 1999, Song and Croft proposed the Good-Turing estimation method to computing language model parameters (Song and Croft, 1999). Rather than using maximum likelihood estimation, they also suggest combining a unigram model with a bigram model through linear interpolation, i.e. using the probabilities of low order n-gram (unigram) to adjust the probabilities of higher order ones (bigram). Their experiment results have shown the improvement over the combination model is marginal. For basic unigram language model, a Bayesian extension provides the full predictive distribution in a form amenable to implementation by traditional IR models (Zaragoza et al., 2003).

As an extension of probabilistic retrieval models, the statistical language model for information retrieval retains the basic theoretical foundations of earlier retrieval methods. Thus, improvements have also involved understanding the formal structure of the language model and comparing it to existing probabilistic approaches, e.g. the structure of the surrounding corpus has been considered in the article written by Kurland (Kurland and Lee, 2004) and a risk minimization structure based on Bayes theory has also been reported (Lafferty and Zhai, 2001).

Successful experiments based on language model approach have also been discussed, e.g. cross-lingual information retrieval (Xu et al., 2001; Lavrenko et al., 2002), distributed information retrieval (Xu and Croft, 1999; Si et al., 2002). Modelling redundancy has also been considered by Zhang (Zhang et al., 2002), who proposed a 3 components redundancy model. Along similar lines, document similarity can be measured in an asymmetric way (Kurland and Lee, 2005).

2.2.2 Data sparseness and smoothing

Language model parameter estimates are usually accomplished by maximum likelihood estimation (Akaike, 1973), where the probability of sequence words is equal to the product value of occurrence probability of every words. One challenge is data sparseness problem, where an unseen word in document would get a zero probability, making the probability of entire queries containing the unseen word equal to zero (Chen and Goodman, 1996; Song and Croft, 1999). It may be caused in two aspects:

On one hand, because the estimator attempts to maximise representing the data, it may "over-estimate". It assigns a zero probability to unseen words ($P(q_i|D)=0$), where the score of the document equates to the product of every individual term probability, so even if other search keywords occurred in the document, the entire probability will equate to zero $(P(Q|D)=0)$.

On the other hand, according to the statistical principle, the accuracy of a statistical estimator relies on the sampling size of the resource. The probability produced by insufficient sampling (such as an individual document) could not be truly trusted. The unseen word would likely occur in larger sampling, e.g. sampling the entire document corpus. Even the estimated probability of a seen word is poor, since they partly occur by chance.

Therefore, the accuracy weakness is compensated for by using smoothing to adjust estimated results. The principle of smoothing is to make the probability distribution more uniform, by adjusting low probabilities or zero probabilities upward, and high probabilities downward. It not only eliminates zero probability, but the smoothing techniques can also improve the accuracy of the model parameter estimation. In the following paragraphs, two different kinds of smoothing approaches will be discussed, i.e. global smoothing and local smoothing.

Traditional global language model smoothing

Zhai introduced a Dirichlet prior method that produced the best smoothing results on the title queries (short queries) searching compared with other traditional global smoothing approaches. For estimating the probability of unseen terms, the Dirichlet smoothing method adopted an overall corpus probability distribution to assign more reasonable probability to unseen words. Let *p (w|Coll)* denotes the entire collection language model. It can be calculated according to the counts of the terms in the entire collection, i.e.

$$\mathrm{p(w|Coll)} = \frac{\sum_{D \in C} c(w, D)}{\sum_{D \in C} |D|} \tag{6}$$

The Dirichlet method imposes entire collection probability prior to estimating the probability of unseen terms in the document. Dirichlet is a conjugate prior for multinomial distribution, which essentially means that the prior (the multinomial distribution of the entire collection) is similar with the distribution of the specific document, thus allowing conversion of the prior into pseudo probability to the sparseness term. The function of language model can be represented as follows:

$$\mathrm{p(w|D)} = \frac{\mathrm{c(w, D)} + \mu \mathrm{p(w|Coll)}}{|\mathrm{D}| + \mu} \tag{7}$$

Where, the μ is a document dependent coefficient which can maximize the probability of unseen words. It can be seen as the extended context for the document without data sparseness and follows the multinomial distribution of the entire collection. The $|D|$ is the length of the document.

The Dirichlet smoothing method would provide extra pseudo counts for all unseen terms according to their overall counts in the entire collection. The more the word occurs in the whole collection, a higher pseudo count will be assigned to the unseen word. As a result, after pooling these pseudo counts with the initial counts of the words observed in the document, it would effectively combine the entire probability with the original term frequency.

Enhanced local language model smoothing

The global smoothing approach estimates the probability of the unseen word based on the same background, i.e. collection language model. Such approaches are unable to discriminate the content of the data; the language model may only indicate the data's structure rather than distinguish data differences.

To overcome this problem, several recent studies show that document corpus structures can be exploited to provide hidden relationships between documents. The basic idea of such local smoothing strategy is to smooth the document based on document similarity or its located cluster/classification (Zhai and Lafferty, 2002; Kurland and Lee, 2004; Liu and Croft, 2004; Mei et al., 2008; Zhou, 2008b). Compared to the global smoothing method, a local smoothing appied document similarity provides "customized" smoothing for each individual document. The local smoothing strategies demonstrated robust effectiveness in several recent related researches, and it is the best smoothing framework so far. There is an important assumption made in such local smoothing, i.e. "documents in the same category should have similar representation". The smoothing performance is directly determined by the adopted approach to model potential relationship between the documents. Liu and Croft first introduced the cluster based language model smoothing (Liu and Croft, 2004):

$$\hat{P}(w|D) = \lambda P_{ML}(w|D) + (1-\lambda)[\beta P_{ML}(w|Cluster) + (1-\beta)P_{ML}(w|Coll)] \qquad (8)$$

Where, cluster language model $P_{ML}(w|Cluster)$ is estimated by the maximum likelihood estimator (ML); β is the coefficient to control and normalise the cluster language model smoothing.

The system first smooths the cluster as a long document, then smooths individual documents based on the smoothed cluster language model. Therefore, could a semantic approach provide more discrimination between documents, to improve the SLM smoothing process? For this initial research into the semantic enhanced IR system, the local smoothing framework was selected attempt to improve the performance of IR system.

2.3 Ontology and Semantic web

2.3.1 Ontology

As previously mentioned in section 2.1, the traditional search engine, cannot provide potentially relevant documents beyond the keywords. This view is reflected in the following two aspects (Xu et al., 2008). Firstly, the retrieval model lacks discretion, i.e. the computer cannot match queries by discretionary judgement (reasoning) over Web content. Secondly, as yet, there are insufficient intelligent search tools for search engine "learning" of Web information.

As the result, since the Semantic Web was proposed in the late 1990's, Ontology has been put forward as one of the potential improvement approaches to overcome the limitation of keyword-based traditional search engines (Uwe M. Borghoff, 1998; Vallet et al., 2005). The fundamental principle of a computing ontology is formalised representation of knowledge agreed for sharing, in a language that provides a logical view of a subject area of domain. Ontology defines concepts (classes), relationships (properties) between those concepts and their constraints (restrictions). According the principals of Ontology, it can be defined in two perspectives (Chandrasekaran et al., 2002):

1. Ontology is a representation vocabulary, often specialized to some domain or subject area.

2. Ontology describes a body of knowledge, which is typically a consensual knowledge domain, which can be learned by computers automatically, e.g. car is a

kind of vehicle, in ontology car class can be specified as a subclass of class vehicle. In other words, the representation vocabulary provides a set of terms used to describe "accepted facts" about some domain of interest.

A simply definition of the ontology has been provided by Gruber (Gruber, 1993; Studer et al., 1998): "*an Ontology is a formal, explicit specification of a shared conceptualisation*". *Conceptualisation* refers to an abstraction of specified reality, in which the basic concepts are identified; *formal* refers to the property that ontology structure is computer readable, e.g. generated by OWL; *shared* means that ontologies capture the consensual knowledge, accepted by people - which corresponds with natural language representation.

The pivotal question is how to use ontological techniques to represent the reality of knowledge to the computer. Fig 4 represents the relationship between knowledge in reality, conceptualisation and a language representation (e.g. ontology) of this conceptualisation, i.e. so called Ullmann's triangle (Ullmann, 1972).

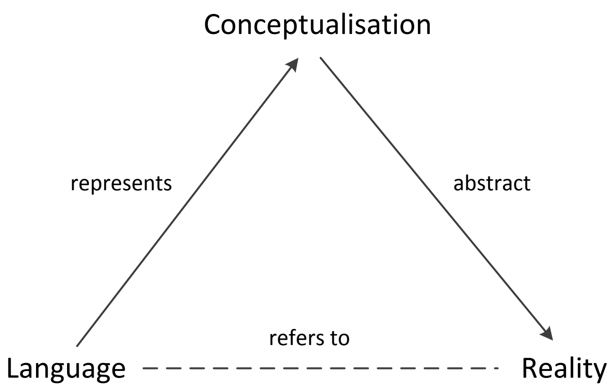


Fig. 4 Ullmann's Triangle

The dashed line between language and reality indicates that language can be used to describe the reality: specific language represents the conceptualisation, which is used to construct the abstraction of reality. The accuracy of representing conceptualization in reality is directly determined by its "description" capability.

Fig 5 was proposed by Guizzardi (Guizzardi, 2007) to describe the relations between reality conceptualisation, ontologies, logical models and intended models. Where, a logical model presents a basic optimal vocabulary which could clearly represent a specific reality, whilst an intended model is a reader/computer's understanding based on the description. As shown in Fig 5, there is a gap between ideal logical model and actual intended model, which easily cause ambiguity communication, i.e. insufficient information from the language represented.

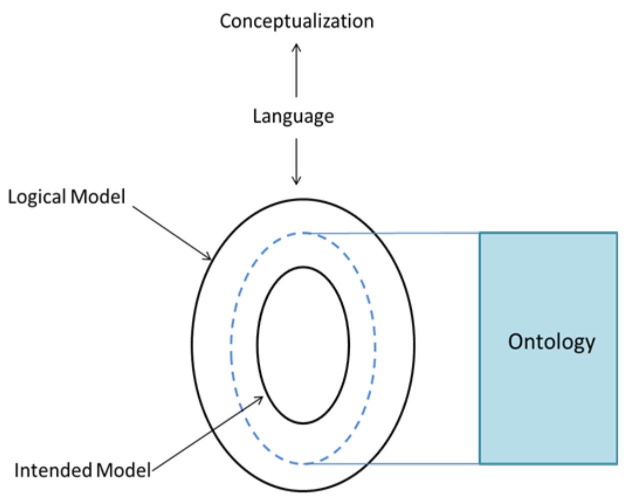


Fig. 5 Relations between language, conceptualisation, ontology

Ontology as a vocabulary between these two models links two models, i.e. concepts and relationships in an ontology are built according to natural language (reality) and the ontology concepts, relevant in some specific domain, are used to describe the resources. In order to reduce this gap to achieve a better communication about reality, Borgo (Borgo et al., 2002) proposed two possible ways to more accurately describe reality:

1. Developing richer language axiomatisation.

Changing the descriptions of reality by more axiomatisation to reduce ambiguity/clarify meaning, i.e. standardization of natural language. Reflect as a reduced distance between ontology models and intended models. However, because of natural language complexity, it is hard to provide a strict standard for it.

2. Enrich domain concepts and relevant conceptual relations.

In principle, it is possible to describe a complex conceptualisation about reality, by extending the knowledge domain and concepts involved by such domain, i.e. more clear description about the concepts. The practice is to use more related, specific terms to describe one domain. This provides a potential way to improve distinction of language model, i.e. provide more computer-understandable information (semantics) about a query and documents to improve information understanding.

2.3.2 Ontology description approach

The basic component elements of ontology are concepts and their corresponding relationships (Paralic, 2003). Concepts can be represented as set of nodes in the ontology graph and usually have a textual description. Ontologies allow a formal definition using some kind of logic, e.g. depicting hierarchies or unions of classes.

A concept may have more than one term to describing it (Andreou, 2005), e.g. synonym “subway”, “underground”, “tube” and “metro” can all used to describe the concept of underground, although "Subway" also refers to the *instant* (name) of a food store (polysemy). Because of natural language versatility, ontologies usually contain several terms for each concept (synonymy) - Furnas et al. (Furnas et al., 1987) proposed that people who have common sense to use the same term to describe the same concept is less than 20%.

Relationships are usually used to describe the properties of two or more concepts (Andreou, 2005). Most ontologies include "is-a (subclass)" and "is-part-of" relationships between concepts, e.g. "Car is-a vehicle", "Shoulder is-part-of human body".

2.3.3 Type of Ontology

The main types of ontologies can be categorized at three levels, as in Fig 6: top-level ontologies, domain and task ontologies, and application ontologies (Guarino, 1998). The degree of ontology complexity can be achieved according to the way they formalize the concepts (Gruber, 1993). A top-level ontology typically represents generalized concepts, as a result it becomes more shareable (applicable) to a wider range of domains and applications but provides limited domain expressivity. Conversely, a fine-grained ontology represents a more specific knowledge that may provide more domain reasoning capability but be less shareable.

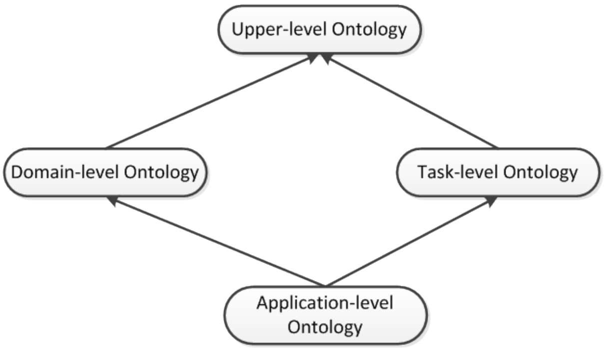


Fig. 6 Ontology type classification(Guarino, 1998)

Top-level ontologies tend to describe abstract general concept terms and relationships like space, time, matter, objects and events, and are domain-independent. Examples are WordNet (Miller, 1995), a lexical database resource for natural language processing systems and the knowledge or commonsense-based Cyc ontology (Lenat, 1995; Noy and Hafner, 1997).

A domain ontology (a more modular ontology) provides a vocabulary about concepts in one specific domain, e.g. medical, engineering or technology. Task ontologies provided for generic task, e.g. market management.

Application ontologies define both domain and task, e.g. related to ticket booking by a specific company. For this initial research, modular ontology was selected as it controls the redundancy during the process.

2.3.4 Semantic web

"The Semantic Web is a web of data, in some ways like a global database. It is an extension of the current Web in which information is given well-defined meaning" (Berners-Lee et al.,2001), in the seminal article written by Tim Berners-Lee, he

initially said that the World Wide Web should provide greater information definition and he emphasized machine friendly knowledge representation, enabling computers to better understand and manipulate information semantics.

The basic theory of Semantic Web is to combine a hierarchy of languages (layers) with current World Wide Web - that enables advanced automatic processing of Web content, i.e. enabled to be collected and processed by both human and computer (W3C, 2004a). The "layer cake" (Berners-Lee et al., 2001) in Fig 7, illustrates a layered structure that conceptualises the various components of Semantic Web:

The Semantic Web is an extension of the current Web - it should be represented as a Web of data supported by a general model of the core Web standards and technologies (Matthews, 2005), to allow information to be shared and reused between different networks and users (Berners-Lee, 2006).

The three bottom layers are the technologies that are applied to existing hypertext Web and that provide the platform for the Semantic Web. The three top layers are mechanisms to represent technologies used to realize Semantic Web capability. The descriptions for these layers are shown in Appendix A.

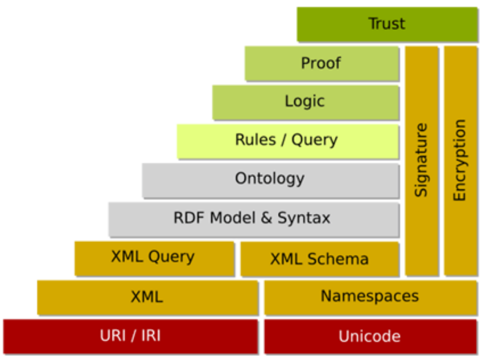


Fig. 7 Semantic Web architecture (Berners-Lee et al., 2001)

The middle of the "layer cake" contains several pivotal Semantic Web language technologies, i.e. the RDF (Resource Description Framework) core language (W3C, 2004c), the RDF Schema (W3C, 2004d) and the Web Ontology languages (W3C, 2004b). They provide a standardized platform to building Semantic Web applications, e.g. distribute current RDF/XML data collection to support data integration and reuse.

RDF & RDF Schema

The Semantic Web aims increase the amount of machine process-able data it retains. Resource Description Framework (W3C, 2004c) is the first enabling layer of the Semantic Web, providing a simple metadata representation framework for Web-based resources (Decker et al., 2000). Whilst RDF syntactic representation is based on XML type syntax, RDF can identify Web-based resources using a directed labelled graph model for describing relationships between the resources.

The underlying structure of any resource description, expressed in RDF format, is a collection of triple structures, where each triple is called a RDF graph, consisting of a subject (identifies what resource object the triple is describing), a predicate (also called a property or attribute) defines a relationship with an another object (another object or value). The RDF graph is shown in Fig 8.

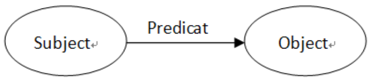


Fig. 8 RDF graph

Thus, the node of a RDF graph indicates subject or object, and the arrow represents the relationship or property between them. RDF structure provides a basic object-attribute-value data model for metadata, in that it is weak in defining how components of a domain are related to each other.

RDF Schema (W3C, 2004b) enables ontology structure to describing classes of resources, range restrictions and properties between them in the basic RDF model. However, The data expressivity of RDF and RDF Schema are limited to some extent, i.e. RDF is weak for multiple predicates, and RDF Schema is weak in various perspectives, e.g. Properties' restriction, Exclusion classes and Reasoning property (Manola et al., 2004),

To overcome the limited RDF framework semantics capability, the W3C identified the Web Ontology Language **(OWL)**, for ontologies in the Semantic Web, which would provide more expressiveness than RDF and RDF Schema.

OWL (OWL Web Ontology Language) is W3C standard recommendation for Ontology representation syntax in the Semantic Web (W3C, 2004b). Ideally, an OWL ontology is represented by RDF graphs, in the sense that OWL is built upon RDF Schema to define the classes and properties (rdfs:property, rdfs: subclassof, etc.). At the same time, it was also derived from the DAML+OIL logic-based ontology language primitives (Horrocks et al., 2002) to improve richer expressiveness identified above.

Different levels of syntax expression and efficient reasoning support prompted W3C to define OWL as three different sublanguages, i.e. OWL Lite, OWL DL and OWL Full.

OWL Full is the entire expression language and contains all language primitives. It is fully upward compatible with lower layer of Semantic Web, i.e. RDF and RDF Schema. It is flexible, i.e. where data/information expression is more important than logicality, which means it cannot be generated by logical reasoning tools automatically.

OWL DL was designed to regain computational efficiency by using a reasonable fragment of First order Logic called Description Logic. It is sublanguage of OWL Full which contains all OWL language constructs with restriction. It can be generated with logic reasoning, but loses full compatibility with RDF: an RDF document has to be extended before it used in a legal OWL DL document.

OWL Lite is useful for data classification hierarchy and simple constraint features, however, it excludes enumerated classes, some (e.g. disjoint) statements, and reasoning properties. It syntactically simplest sublanguage of them; thus it is easier to grasp and easier to implement.

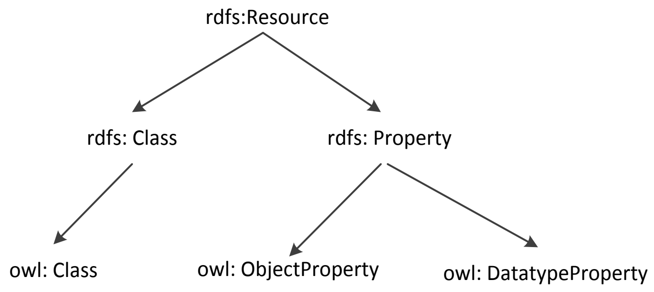


Fig. 9 Relations among the RDF, RDF Schema and OWL (Antoniou and Harmelen, 2009)

Fig 9 demonstrates the relationships between the RDF, RDF Schema and OWL, i.e. all varieties of OWL according to RDF/RDF Schema syntax, instances of ontology are declared in RDFs and OWL constructors are all extended by their corresponding RDF part (triple structure).

2.3.5 Semantic web applications

More recently, Semantic Web agents and other automated processes have produced more information faster and efficiently using Semantic technology-based applications, especially in the following areas:

Search Semantic Web ontologies

Ontologies of the Semantic Web are represented by using Web Ontology Language (OWL), which explicitly represents the meaning of Web content terms and the relationships between those terms (Bechhofer et al., 2004). As a unique identifier for Web entities and relationships, URI references (Berners-Lee et al., 1998) are the basic resource identifiers of Semantic Web ontologies. In the Semantic Web, Ontology creators/developers need to search ontologies for extension, combination or other proposes. For a huge "relationship" network, the information publisher may search popular ontologies to develop their information domain. E.g. the FOAF (Friend-Of-A-Friend) ontology (Brickley and Miller, 2005) is an application for creating a linked person framework Web of computer-readable pages, i.e. to provide links between them according to their self-developed profile descriptions. Swoogle (Ding et al., 2004) is a search engine for Semantic Web ontologies, where ontology engineers can search the ontologies for further processing information.

Updating existing ontologies

Semantic Web ontologies are often interlinked, but they evolve independently. Prior to updating the existing ontologies, ontology developers not only need to increasingly update the new concepts into existing ontologies, but also need to identify and evaluate potential mistakes on those ontologies. E.g. Pellet (Horridge and Tsarkov, 2006; Tsarkov and Horrocks, 2006) is an efficient description logic reasoner compatible with OWL-DL, which is used to compute and validate the internal logic relationship between ontological entities. Interestingly, the first Protégé ontology was imported by 38 ontologies in 1987, whereas it has been updated more than 600 times during 2013.

Enumerating relevant document

In a celebrated article it was stated that “The Semantic Web is a web of data, in some ways like a global database”(Berners-Lee, 1998), Semantic web based applications consider the Semantic Web as worldwide database. It requires every Semantic Web document to use a specified namespace to uniquely identify and disambiguate content. In this huge web database, users can conduct information retrieval or execute searches for particular instances or concepts, e.g. The Semantic Web based applications for the petroleum industry (IFP) (Braunschweig and Rainaud, 2005) and market management (Maedche and Staab, 2005) all treat Web hypertext and related components as the items of a database. FOAF personal curriculum vitae have been found in a large numbers, and have been used as a database in Semantic Web based applications i.e. SECO[1].

Compared with traditional commercial search engines, Google, Yahoo and Bing, the Semantic Web embedded search engine has a better capability to identify and present Web data concepts and their relationships; thus it may offer a valuable benefit to solving the data sparseness problem in SLM base IR systems.

2.4 Query expansion

Natural language ambiguity originates from the distance between a logic description model and an intended description model; as a consequence a search query generally results in a long list of documents being returned. It is difficult to use limited query terms to accurately represent an information requirement. To improve the accuracy of

[1] http://ecat.secotools.com/

the retrieval system, queries need to be disambiguated by applying context into an information retrieval system. Query expansion techniques rely on contextual information to resolve ambiguities. Contextual information can be acquired from relevance feedback, Corpus-dependent knowledge (e.g. term occurrence frequency), and Corpus-independent knowledge (e.g. ontology based query expansion). This chapter will concentrate on these three different query expansion aspects.

2.4.1 Overview

Query expansion is needed to eliminate the ambiguity of natural language and also the difficulty in using limited terms to represent an information concept. The main purpose for employing query expansion in IR systems is to include additional, meaningful terms to the initial query. This approach should provide richer contextual information to better express an information requirement. Much work has been done in the area of query expansion, and has achieved remarkable improvements. Research has observed that accurate query expansion can provide better precision (Mitra et al., 1998) and recall (Krovetz and Croft, 1992).

Number of new terms

It is possible for a query expansion process to generate a large number of indicative terms, relating to the initial query, that it might not be practical to use all of them. Some research has concentrated on the optimum number of terms to include. There are different points of views: 1/3 of the indicate terms (Robertson and Willett, 1993), 20 terms (Harman, 1992), and more than 300 terms were added as query expansion (Buckley et al., 1995; Billerbeck and Zobel, 2004). In 2004, Sihvonen and Vakkari (Sihvonen and Vakkari, 2004) proposed that the type and quality of chosen terms is much more important than the amount of them. In the same year, Billerbeck and Zobel (Billerbeck and Zobel, 2004) systematically evaluated the efficiency of retrieval systems that included term expansion. The experiment results showed the optimal number of query expansion terms is different from query to query. It is a challenge in this project that to identify the optimal number of expanded terms to achieve optimal result.

Weighting of new terms

To resolve the problem of quality of expanded terms, term weighting was applied to the query expansion process. Rocchio (Rocchio, 1971) used the vector space model, where the documents and queries were represented as vector in the information space. The vectors are weighted, which differentiates degree of similarity to the initial query terms. After relevance feedback the weights were adjusted, and the results produced positive results.

Voorhees (Voorhees, 1994) improved the retrieval accuracy by adding a coefficient (between 0 and 1) to each expanded term. In 1990, Robertson proposed a Probability model for IR based on Probability Ranking principle (Robertson, 1990). This approach ranks the document according to the relevance judgement, which is the probability of a document being relevant to a query. The weighting for each term is modified based on the so-called F4.5 formula:

$$w_t = \log \frac{(r + 0.5)(N - n - R + r + 0.5)}{(n - r + 0.5)(R - r + 0.5)} \tag{9}$$

Where, w_t is term weighting

r is number of relevant documents contain the term t

R is total number of relevant documents

n is number of documents contain the term t

N is total number of documents in the collection

2.4.2 Relevance feedback based query expansion

Relevance feedback is a technique for modification of the initial query, using words from document results or identified relevant documents (Salton and McGill, 1983). It is a method to improve the quality of a query term based on a user's judgements.

The query expansion process generate new terms based on the top-ranked documents retrieved, according to the user's initial query. Indicative relevant documents are determined by the user or based on pseudo relevance feedback, where top ranked n documents are assumed to be relevant. Terms from indicative relevant documents are selected to add into initial query.

From the experiments carried out in (Tombros and Sanderson, 1998), it was found that a number of relevant documents in the original document collection directly determine the quality of the expanded query. If the relevant documents are relatively small then the quality expansion can be poor due to the insufficient background information.

Corpus structure can also be used to perform query expansion. Global techniques rely on analysis of a whole collection to discover word relationships, local techniques emphasize analysis of the top-ranked documents retrieved for a query. While local techniques have shown to be more effective than global techniques in general, Xu and Croft (Xu and Croft, 2000) compared two approaches in query expansion and found that local techniques only achieved positive results if all of the top-ranked document are relevant.

Vakkari et al. (Vakkari et al., 2004) compared the performance of interactive query expansion with automatic query expansion. They stated that interactive query expansion achieved better results if all retrieved relevant document were counted. If a user can indicate the relevant term/phrase, relevance feedback is more accurate. Query expansion suggestion not only increases the recall but also interaction enables the user to increase precision.

Ruthven and Lalmas (Ruthven and Lalmas, 2003) found that domain specific collections perform better with relevance feedback than domain independent collections because it is easier to select good expansion terms. The ambiguity of search terms is less significant.

In conclusion, effectiveness of relevance feedback based query expansion can vary depending on many factors. Relevant research provide potential factor to improve the query expansion performance such as corpus structure, user interaction and domain specific background query expansion.

2.4.3 Corpus dependent knowledge based query expansion

Corpus dependent knowledge refers to the corpus contents and potential relationship between terms, e.g. term occurrence frequency and document clustering.

Term co-occurrence based query expansion

Co-occurrence refers to two or more terms situated near/next to each other in the document; it can be interpreted as an indicator of semantic proximity. Chu (Chu et al., 2002) proposed a novel knowledge-based query expansion technique to rewrite/expand a user query. General conceptual terms in a query are substituted by relevant specific terms that co-occur with the initial query concept. Because the expanded specific terms reduced ambiguities, it produced better retrieval effectiveness than unexpanded queries. The drawback of the approach is that specific terms reduce the recall. It is therefore only suited to precision-oriented retrieval.

Vechtomova (Vechtomova et al., 2003) presented two long-span collocates query expansion techniques, i.e. global collocation analysis and local collocation analysis. They use words that significantly co-occur in topic size windows with query terms in the resource documents. Whereas, with global collocation analysis, it captures terms from the entire collection, local collocation analysis only extracts terms from a subset of top ranked documents. The experiment showed that global collocation analysis achieved worse results compared with the low expansion one. This may be because the terms generated from entire collection are too general and results lead to more ambiguity. The local collocation experiments produced better results.

Cui (Cui et al., 2003) described a new term co-occurrence query expansion approach which was based on user interaction recorded in a user log. They take advantage of user judgement, implied in user logs, and expand the initial query according to the extracted correlations between query terms and document. Experimental results revealed that such query expansions produced better retrieval effectiveness.

Concept structure based query expansion

Lexical networks were used to carry out query expansion techniques from early nineties. They contained domain-specific vocabularies and relationships between them. Lexical relationships between terms were utilised to generate new terms. Pustejovsky (Pustejovsky, 1995) proposed a novel query expansion approach to use machine-readable dictionaries and large text corpus, to construct a lexical network and refine initial queries through statistically based corpus acquisition methods.

Concept hierarchies are another approach for deriving additional terms (Joho et al., 2004). The concept hierarchies are automatically generated from the document collection by extracting the salient words from the top ranked documents. The extracted words are hierarchically organised using a subsumption function to determine the relationship between the terms. If one concept subsumed another concept then the former should be placed as parent concept, just as the superclass in the ontology (Sanderson and Lawrie, 2000). The inverse document frequency (idf) is used as term weighting to indicate whether a term was general or specific. The experiment produced positive results.

In conclusion, corpus dependent knowledge based query expansion concentrates on obtaining indicative terms from the document collection. They are reasonable approaches for static document collection. However, for web collections, because they are more dynamic and fluid, the knowledge models would have to be constantly updated and generated repeatedly, which would be a time consuming activity.

2.4.4 Ontology based query expansion

The limitation of relevance feedback techniques and corpus dependent knowledge based query expansion is that they analyse corpus content to extract the additional words, which relies on sufficient relevant documents and also that these documents contain a set of terms that related to the query. Corpus independent knowledge techniques are proposed to avoid these drawbacks.

Corpus independent knowledge techniques can be employed using Ontology. The purpose of an ontology is to provide a computer readable context for the vocabulary disambiguation. Ontologies provide consistent vocabularies to precisely represent the knowledge domain (Gondy et al., 1999). Ontologies range from general to domain-specific types. The following subsection will discuss query expansion using general ontologies and domain specific ontologies respectively.

Domain-independent OQE

WordNet has been a popular general ontology used in query expansion since the nineties. Voorhees (Voorhees, 1993) first utilised WordNet based query expansion to improve retrieval performance. They compared the retrieval performance of WordNet based word sense disambiguation terms and stem terms. The experiment results showed that the effectiveness of retrieval produced by word sense disambiguation was worse than using stem terms. The findings indicate that short query can be difficult to disambiguate by general thesaurus because IS-A hierarchy in WordNet is insufficient to reliably select the correct terms.

Gonzalo (Gonzalo et al., 1998) carried out an experiment to evaluate search performance with three types of query methods: original query, WordNet synsets, and manual disambiguation of term senses. Their results show that if the query is not disambiguated, WordNet synset-based query expansion performs only as good as the original query. Query expansion using automatically generated synsets resulted in missing correct matches which had a deteriorating effect on retrieval performance. Meanwhile, with manual disambiguation of word terms, the system achieved 29% search effectiveness improvement.

Finkelstein et al (Finkelstein et al., 2001) established a context based search system, where a query was disambiguated by deriving the context from the text surrounding the query terms in a given document. The system used a semantic network to measure the distance between the indicative term and initial query. Linguistic information (e.g. hypernym and hyponym) is obtained from the WordNet dictionary. The results showed that using contexts can markedly improve performance.

Jones(Jones et al., 1995) applied the INSPEC thesaurus and used eight relational databases to store the potential relationships between the terms, e.g. equivalence and subsumption hierarchy. Their experiment results suggest that quality of the thesaurus is paramount and there is no correspondence between the number of expanded terms and the query performance. A Thesaurus that has great coverage, depth and accuracy of concepts has an improved chance of achieving better performance (Jones, 1993). They also found that the number of expanded terms should be determined by the distance from the indicative term to the initial query.

Domain specific OQE

The drawback of a general, domain-independent ontology (e.g. WordNet) based query expansion is in the quality of the expanded term, because they have a broad coverage and general, potentially ambiguous terms within the ontology can be problematic. Domain-ontologies can eliminate the problem. The terminology in these ontologies is less ambiguous therefore, it largely narrows the concept coverage and queries can be expanded with a higher chance of accuracy. Domain-specific ontologies have been constructed in many applications in related research.

Fu (Fu et al., 2005) constructed an ontology-based spatial query expansion approach that retrieved documents considered to be spatially relevant. In their work, the query was expanded according to both a domain ontology and a geographical ontology, i.e. spatial terms (e.g. place name) were expanded by the geographical ontology, whereas non-spatial terms (e.g. north-of and inside) were modelled in a tourism domain ontology. They achieved positive search results.

Nilsson (Nilsson et al., 2011) proposed a prototype for ontology-based cross-language information retrieval in a restricted domain. The system is based on SUiS (Stockholm University Information System), which does not allow free-form queries; it restricts the query terms to who, what, where and when. Synonyms and hyponyms in a domain specific ontology are used for query expansion. The experiments have shown an improvement of precision in results.

In the medical area, Bao (Bao et al., 2004) also utilised a domain-specific ontology and domain-independent ontology for colonoscopy video database annotation. The domain-independent ontology contained general information on properties of video, and the domain-specific ontology included the colonoscopy terminology. The experiments have shown an improvement in results.

Díaz-Galiano (Díaz-Galiano et al., 2009) utilised the medical ontology MeSH to improve search performance of a multimodal (textual and visual) information system by expanding a user's query with medical terms. KL-divergence weighting scheme and pseudo-relevance feedback were applied to control the quality of expanded terms. The experiments obtained improved results in both textual IR system and the combined approach (textual plus visual information searching).

In conclusion, advantages offered by ontologies are that they are machine readable. Ontologies offer benefits to corpus-independent knowledge based query expansion and even more useful in specialised information retrieval task. The quality of the ontology directly affects the expansion performance.

General ontologies are more suitable for a general, broad search purpose, because of the wider coverage. However, handling ambiguous terms with an ontology can be problematic. In contrast, a domain-specific ontology describes concepts and terms in a specific area, it narrows the search coverage but increases the chance of query expansion accuracy. Jones and Chen (Chen et al., 1993; Jones, 1993; Jones et al., 1995) suggested that the quality of knowledge model based query expansion can be controlled by term weighting.

2.5 Literature review conclusion

The literature review chapter, conducted a review of relevant research of statistical language model based information retrieval and ontology, as well as studies of query expansion in term of different approaches.

By evaluating principles of IR and the data sparseness problem of a language model, ample evidence has been found that exploiting semantic contexts could be a useful approach to improve search precision and recall. Ontology based query expansion (OQE) could be applied to solve the problem of lack of distinction and word ambiguity to improve search performance. Further, ontology based document classification (ODC) can exploit hidden relationships between documents, to produce a more accurate smoothing process.

However, the Semantic Web community only recently appears to have become fully focused on Semantic search, and there are no significant examples in the public domain. None of the existing work offers a framework to exploit ODC and OQE into language model.

Hypotheses for issues identified

The search experiments will be used to test the following research hypotheses:

i. Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking. This will be tested by comparing search effectiveness of keyword-only query against various OQE enhanced search model – see OQE + entire, in subsection 4.5.1 and chapter 5.

ii. Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. This will be tested by comparing the non-weighted baseline with entropy-weighted baseline and tf-idf-weighted baseline.

iii. Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or classification-based local smoothing strategies are more effective than the global strategies. To consistently test affect by exploit ODC into language model, this hypothesis will be tested by comparing KO+ODCs with KO+entire (baseline).

3. RESEARCH METHODOLOGY

This chapter illustrates the background to the research experimentation approach by showing how a language model could exploit OQE and ODC to improve search effectiveness. The experiments compared 53 different ontology enhanced search modes with the baseline (Keyword-only searching with Dirichlet smoothing) using TREC WT2g document corpus.

The envisaged benefits of ontology enhanced language models could be in improving relevant document ranking (precision) and retrieving a more comprehensive set of relevant documents (recall). Since a typical Web user might be more interested in examining fewer search pages, to find relevant documents, (requiring high precision) and have no real interest in knowing every relevant document is retrieved (high recall), the evaluation of search performance experiment will primarily consider precision outcomes in the low recall intervals, i.e. first three recall intervals (10%, 20% and 30%). The success of the established search model will be determined by comparing the corresponding P&R curves against the baseline.

The methodology for results comparison will involve several activities, including constructing a search model based search process, developing a search interface, identifying programming techniques, ontology construction and extracting corresponding concepts, baseline testing and setting smoothing parameters, and calculating document scoring for ranking.

3.1 SEARCH MODEL COMPARISON PROCESS

Since the objective of this project is to compare the search performance of using OQE and ODC against the basic language model as the baseline. Term weighting is also utilised to control the quality of the query. A control set of query terms will be used for baseline and OQE and/or weighting and/or ODC enhanced search models, i.e. each query comparison will be executed first in baseline, and then the same query term set will be reused for expansion in the various enhanced language models.

OQE and ODC provide two different ways to improve the SLM based document retrieval, i.e. OQE produces more relevant keywords in query term sets and ODC provides a document corpus structure exploited in the smoothing process. Term weighting approaches are assigned to terms and/or added terms to control the quality of them.

For the purpose of separate analyses, the improvement by exploiting OQE and ODC in the language model will be divided into 4 groups - using comparisons between both baseline and enhanced search models. The following flowcharts essentially illustrate high-level views of the key steps in a typical language model search process. Different combinations of query processes, weighting processes and document processes are denoted in red.

Evaluate OQE enhanced SLM-based IR performance

Search performance evaluation will be based on a comparison between baseline (a) and OQE-only enhanced search model (b):

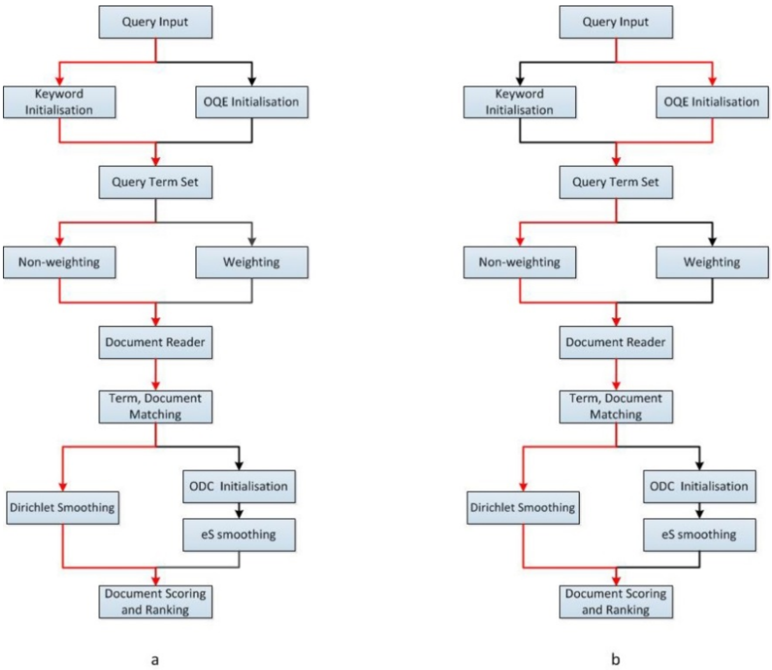


Fig. 10 High-level search process of baseline and OQE enhanced model

a: Baseline of the experiments - search for user's input keyword-only and smoothing the estimated document model using Dirichlet approach according to the background information of entire collection.

b: OQE-only enhanced search model - search for expanded query terms against document corpus and utilised the same smoothing approach to baseline.

Evaluate ODC enhanced SLM-based IR performance

A comparison will be produced between baseline (a) and ODC-only enhanced search model (c):

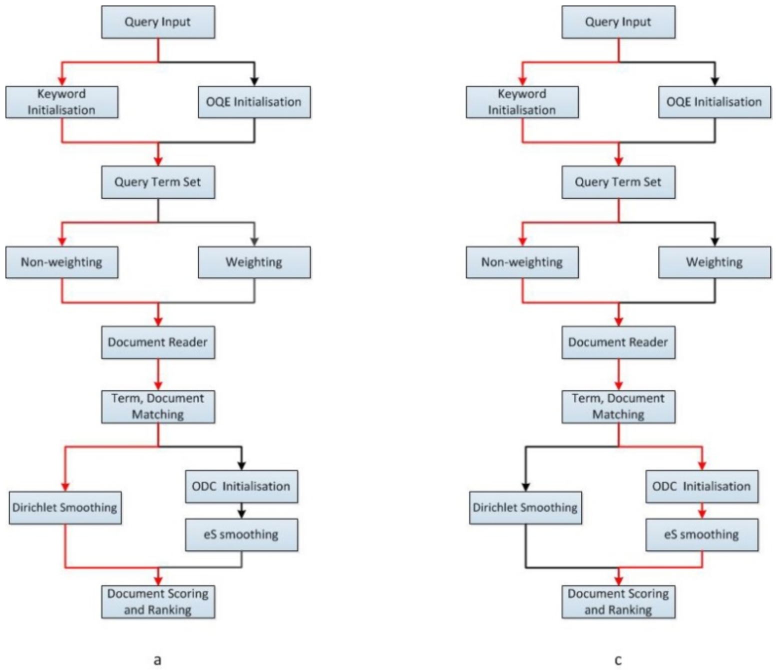


Fig. 11 High-level search process of baseline and ODC enhanced model

c: ODC-only enhanced model. Instead of using entire collection for global smoothing, the ODC will be used for local smoothing to provide more accurate parameter.

Evaluate Weighting enhanced SLM-based IR performance

There are two sets of comparisons in this group:

1. Baseline (a) versus Weighting-only enhanced search model (d). The comparison produces an evaluation for adding weighting to the initial query.

2. OQE enhanced search model (b) versus OQE with weighting enhanced search model (e). An evaluation of weighting effect is constructed by comparison

between weighted OQE and non-weighted OQE.

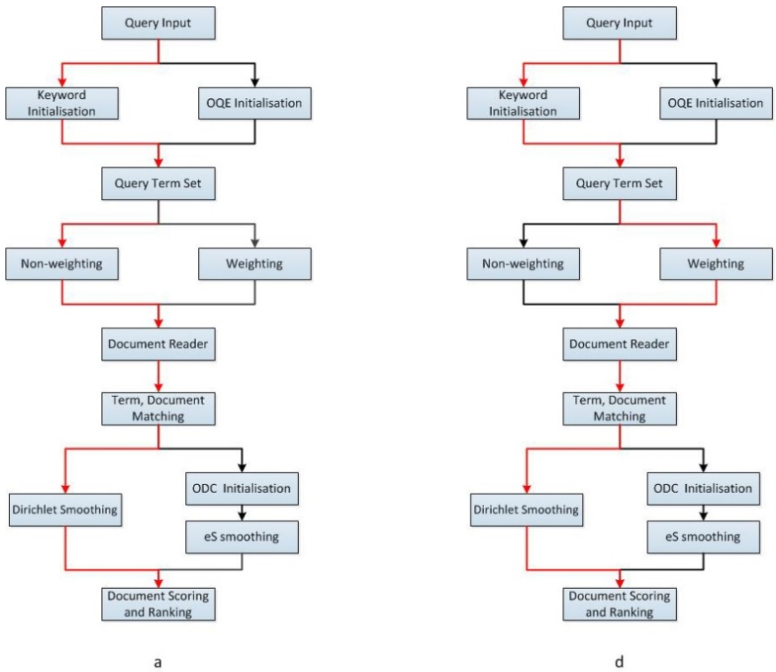


Fig. 12 High-level search process of baseline and weighting enhanced model

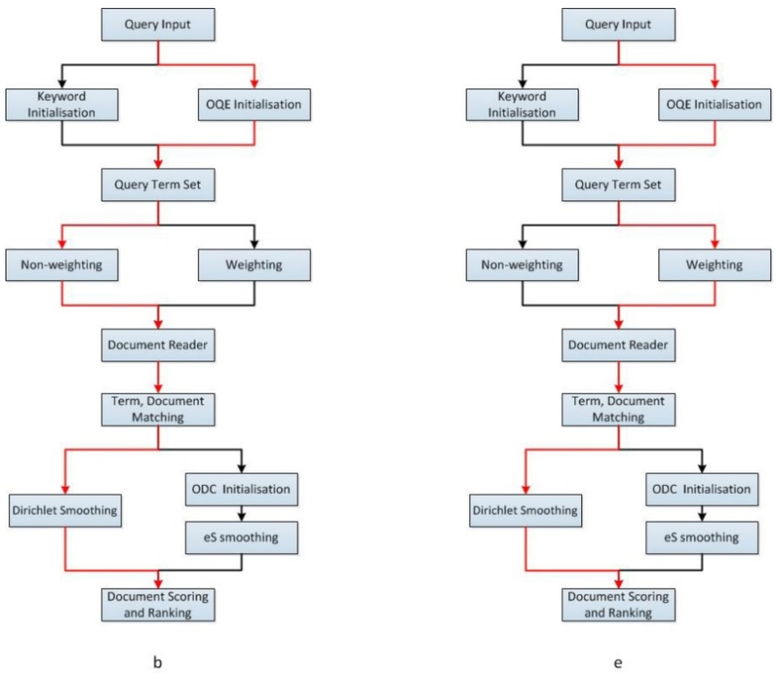


Fig. 13 High-level search process of OQE and OQE with weighting enhanced model

d: Weighting-only enhanced model. Corresponding weightings are assigned to the initial query to control the quality of queries.

e: OQE with weighting enhanced model. Term weightings are used to reduce the potential effect of expanding non-accurate terms during the OQE process.

Evaluate other OQE, weighting and ODC conjugated search model

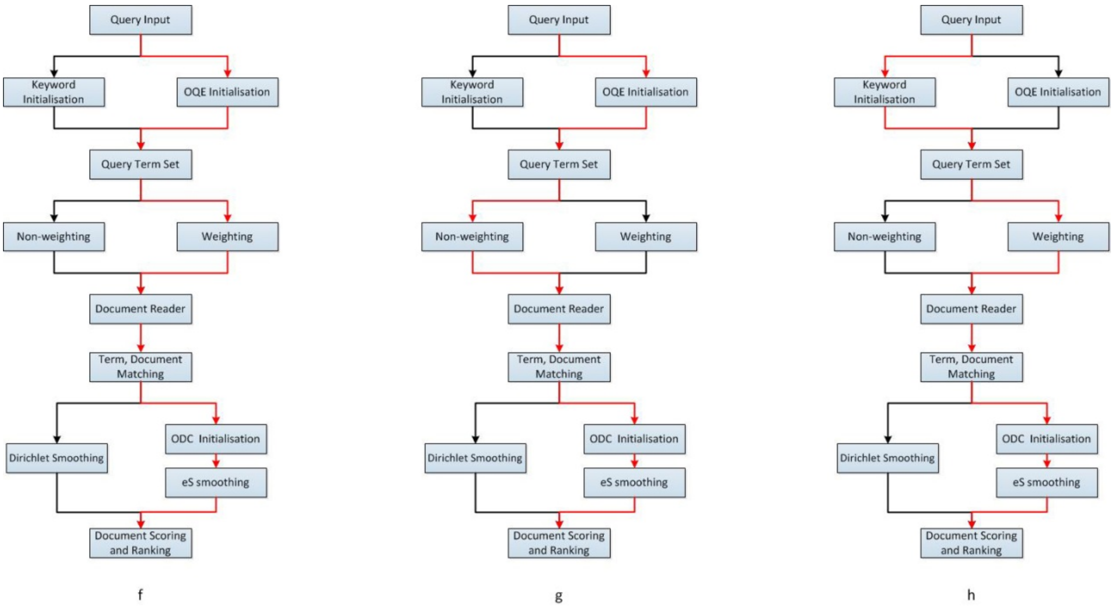


Fig. 14 High-level search process of OQE and ODC conjugated search model

f: OQE + ODC with term weighting enhanced model. This is the most complex search process in all approaches, where OQE with term weighting is merged with ODC.

g: OQE + ODC without term weighting enhanced model. This is the combination of OQE and ODC processes.

h: ODC-only with weighting enhanced model. The system uses a weighted initial query to capture the document and smoothing by ODC enhanced local smoothing.

3.2 DESIGN AND DEVELOPMENT OF SEARCH TOOL LMST

To provide independent experiment control, traditional search engines will not be used and a Language Model Search Tool (LMST) will be developed to facilitate the experiments. The development of LMST will be on the basis of George's semantic search tool (George, 2010) framework, which was constructed to combine vector space model document relevance ranking with Semantic Web technologies, to execute

OWL-based OQE in searches. The LMST similarly employed the same Jena Ontology API methods, to traverse OWL ontologies and extract concepts for OQE and ODC.

The primary purpose of LMST is to provide a prototype search tool to support search experiments:

A query process: employs different query expansion options, e.g. sub class based OQE **(SUB)** or sub & super classes base OQE **(SUPER).** The search tool will also need to provide flexibility in setting the expanded term weighting condition, e.g. **Entropy** based weighting or **tf-idf** base weighting.

A document process: uses various document classification options, e.g. baseclass ODC **(base)**, subclasses ODC **(sub)**, superclasses ODC **(super)**, class hierarchy structure based ODC **(separate)** or class hierarchy structure with term frequency based ODC **(separate_FC)**.

To ensure reliability and validity of the LMST search process, a small test document corpus was created to provide a control set having pre-determined outcomes. The testing was first conducted at each stage of the search progress to verify the accuracy of algorithm developed for: expanding relevant queries, identifying potentially relevant documents, term weighting, generating document categories and P&R statistics for search effectiveness evaluation. After initial testing was satisfactorily validated, the LMST was used to conduct formal search experiments.

3.3 ONTOLOGY SPECIFICATION TO SUPPORT OQE AND ODC

Modular ontology is based on a formal vocabulary of a specific domain, i.e. makes the assumption that keyword expansion can be achieved by a user selecting a topic relevant ontology. Enhanced correlations between the user's input keyword and the ontology hierarchy should provide straightforward relationship between the queries (OQE) and potential relationship between documents (ODC). Base on a potential user's OQE and ODC selections (section 5.1), OQE and ODC processes need to control the range of expansion, i.e. subclasses and superclasses for OQE, together with baseclasses, subclasses, superclasses and whole ontology for ODC. A further requirement is to be able to prevent overload of the OQE process, which could result in deviating too much from original non-expanded concepts. Therefore, different relevance weighting will be considered in the OQE process to reduce the effect of expanded term, i.e. entropy based weighting and tf-idf weighting.

In order to consistently evaluate search performance, there are some bespoke ontologies would developed for 5 different topic experiments. The modular ontology employed was a refinement on the existing ontologies (section 5.2). Prior to the experimentation stage, some trial ontologies were developed for the ontology traversal tests. Protégé was used to develop all ontologies in OWL DL format.

3.4 SMOOTHING PARAMETER SETTING

Dirichlet smoothing performs best in simple global interpolation smoothing with short queries (Zhai and Lafferty, 2002), therefore it was chosen as the baseline for experiments. The standard approach for determining the Dirichlet smoothing parameter is from training data which also contains independent queries and relevance judgments. Because of insufficient resources with topic and relevant judgements, comparison experiments selected a μ value from 1000 to 100,000 to maximize the search effectiveness (i.e. to obtain the highest precision from "training" data). The search effectiveness measurement will be based on potential optimized results for baseline and established search models.

3.5 BASELINE TESTING

As the generation of modular ontologies for topic searching experiments proved time consuming, only 5 topics were selected to evaluate ontology enhanced search models. Search processing utilised the optimal parameters μ for both baseline and established search models. To ensure consistency of the un-trained baseline, baseline was tested using LMST and compared with other research results, i.e. using same queries which are list in Appendix D, same document corpus (WT2g), same searching model (SLM) and same smoothing approach (Dirichlet). The baseline test results listed in Appendix E demonstrate that the optimized baseline of this project is slightly higher than other researcher's trained-parameter searching work, i.e. 35.28% versus 34.4% at the first 30% recall and 18.33% versus 17.63% for overall recall (Blanco and Barreiro, 2008; Lv and Zhai, 2009), i.e. the results are consistent with the precisions identified by other researchers.

3.6 CALCULATION OF SCORING FOR RANKED DOCUMENT LIST

The retrieval experiment will require an independently verified set of documents, i.e. a text document collection with predetermined queries and identified relevance judgements. In 1999, TREC-8 (Voorhees, 1999) was built for "small Web" retrieval tasks, 247,491 documents were assembled as the WT2g collection. WT2g comprised of a set of 50 topics that were each supported by a topic statement and relevance judgements which listing a pool of relevant and irrelevant documents distributed randomly across the full document collection.

The matching between each of the query terms against each document required a mechanism to store the language model algorithm components, e.g. term frequency, document length and term frequency in the entire collection until the complete document corpus had been interrogated. Search calculations could be handled by storing the algorithm components in corresponding array lists, so that the data could then be used in scoring algorithm, to derive ranked weighted documents.

4. MODULAR ONTOLOGY ENHANCED STATISTICAL LANGUAGE MODEL

A robust retrieval system relies on the accurate representation and matching between the user's information requirement and documents. To more accurately represent the query and documents, in this research, we propose ontology based query expansion (OQE) for query disambiguation, and document classification (ODC) to solve the problem of insufficient sampling of documents.

This chapter will demonstrate the relevant techniques used to support ontology enhanced statistical language models. This will include ontology construction & traversal, OQE, OQC and formulas of establish models.

4.1 Ontology Context

Modular ontology was utilised for OQE and ODC process, such that it is built in a modular manner, i.e. based on the context topic. In this research, the modular ontology is built for each selected topic; these were Foreign Minorities, Bone Health, Three Gorges Project, Robotic and Tourism. Some of the ontology context were manually refined from relevant exiting ontology and conceptualised according to the query topic statement narrative and relevant Web site page content. There was no prior reference to the actual TREC corpus documents themselves (See section 5.2)

Ontologies were firstly generated using the Protégé ontology editor and then validated with a description logic reasoner (Pellet). Pellet provides semantic reasoning to support the Protégé classification and inferencing process, i.e. it analyses and validates the ontology hierarchy, identifies any OWL syntax inconsistencies, and verifies any changes to the ontology specification. Ontology consistency is the key to ensuring accurate ontology traversal during the OQE and ODC process.

Fig 15 demonstrates a reasoning process: every circle represents an ontology concepts and relationships between the concepts are represented by arrows (initially, manually developed using Protégé). Stage *a* illustrates an un-reasoned ontology, in which equivalent classes are represented in orange circles, i.e. D5 $\equiv$ D4 and D2 $\equiv$ D1. However it illustrates an inaccurate relationship between concepts, e.g. only D2 is the subclass of D, even though D1 and D2 are equal.

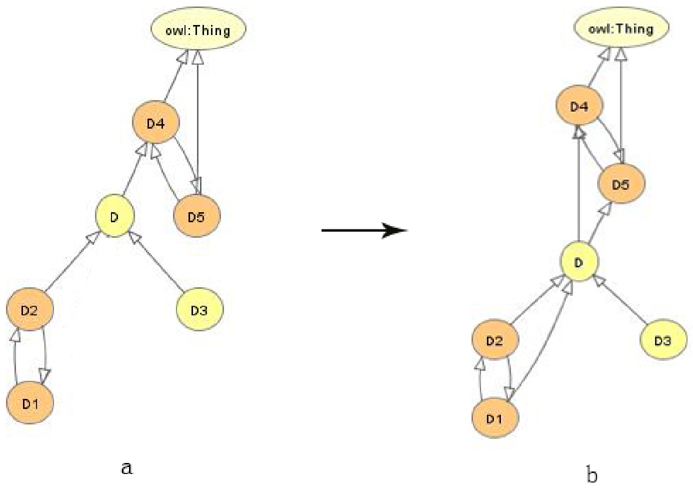


Fig. 15 Ontology concept reasoning process

```
<rdf:RDF
xmlns:owl="http://www.owl.com/example.owl#">
  <owl:Class rdf:ID="D2">
     <rdfs:subClassOf rdf:ID="D"/>
     <owl:equivalentClass rdf:resource ="#D1"/>
  </owl:Class>
  <owl:Class rdf:ID="D1">
     <owl:equivalentClass rdf:resource="#D2"/>
  </owl:Class>
  <owl:Class rdf:ID="D4">
     <owl:equivalentClass rdf:resource ="#D5"/>
  </owl:Class>
  <owl:Class rdf:ID="D3">
     <rdfs:subClassOf rdf:resource ="#D"/>
  </owl:Class>
  <owl:Class rdf:ID="D">
     <rdfs:subClassOf rdf:resource="#D4"/>
  </owl:Class>
  <owl:Class rdf:ID="D5">
     <owl:equivalentClass rdf:resource="#D4"/>
  </owl:Class>
</rdf:RDF>
```

a

```
<rdf:RDF
xmlns:owl="http://www.owl.com/example.owl#">
  <owl:Class rdf:ID="D2">
     <rdfs:subClassOf rdf:ID="D"/>
     <owl:equivalentClass rdf:resource ="#D1"/>
  </owl:Class>
  <owl:Class rdf:ID="D1">
     <rdfs:subClassOf rdf:ID="D"/>
     <owl:equivalentClass rdf:resource="#D2"/>
  </owl:Class>
  <owl:Class rdf:ID="D4">
     <owl:equivalentClass rdf:resource ="#D5"/>
  </owl:Class>
  <owl:Class rdf:ID="D3">
     <rdfs:subClassOf rdf:resource ="#D"/>
  </owl:Class>
  <owl:Class rdf:about="D">
     <rdfs:subClassOf rdf:resource="#D4"/>
     <rdfs:subClassOf rdf:resource="#D5"/>
  </owl:Class>
  <owl:Class rdf:about="D5">
     <owl:equivalentClass rdf:resource="#D4"/>
</owl:Class>
</rdf:RDF>
```

b

Fig. 16 OWL syntax at specification stage a and b.

Following some search experiment tests, it was established that, as LMST uses Jena toolkit libraries, the ontology traversal process also interpreted the same problem, as shown in Fig 16 part a, where a search for the subclass of D did not identify D1.

This ontology traversal problem was resolved by verifying the ontology using the Pellet tool. The results are shown in stage b, where inference has classified the full relationship between D1 and D, and D with D5. The corresponding modified OWL syntaxes are provided in Fig 16 part b, which shows the effect of applying the reasoner to infer ontology classes.

4.2 Ontology based Query Expansion

Statistic language model based semantic search objective is to identify contextually relevant OQE-based terms for disambiguation and increase term sampling in document to accurately estimate the probabilistic distribution of the document. This subsection will demonstrate the ontology traversal paths executed in LMST's OQE process through an example ontology in Fig 17. The ontology contains various concepts in a tree, the names of the concepts require no particular meaning themselves.

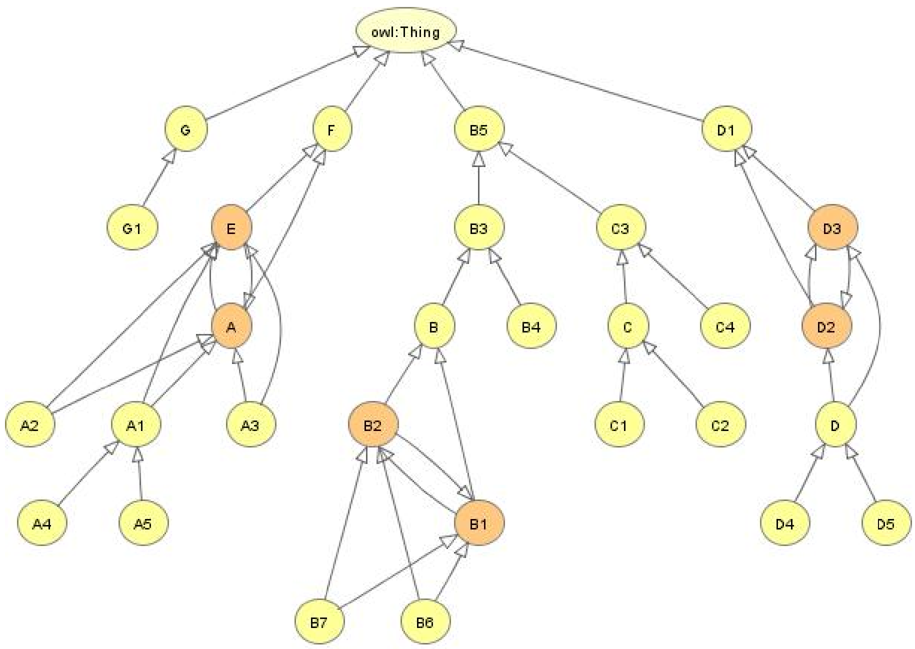


Fig. 17 Ontology example for concepts A, B, C and D

For this example of subsection, it is assumed that all input keywords can be matching to the ontology classes, i.e. four user's input keyword A, B, C and D are matching to an ontology class, so called keyword-matched class (KMC) with potential subclasses, superclasses or equivalent-classes.

The LMST OQE process creates a query term (QT) set for each KMC, i.e. expanding the user's input keyword by traversing the ontology class hierarchy and adding sub, super and equivalent classes to the QT set, based on the search mode option required. There are two OQE approaches established in this research:

(i) SUB OQE, implies adding classes explicitly specified as equivalent and subclass of each KMC. The goal of collecting KMC subclass set is to improve precision and/or recall without introducing heterogeneity. Because the subclass subsumed by a KMC inherits the properties from it, anything that is necessarily true of a KMC is also necessarily true of its subclass, i.e. the expanded subclasses of keyword should also satisfied user's search requirement, and provide more characteristics. Table 1 demonstrates the SUB OQE results using the ontology represented in Fig 17.

Table 1 SUB OQE traversal outcomes

KMC	A	B	C	D
OQE Results	KMC: A EquC: E subC: A1 subC: A2 subC: A3 subC: A4 subC: A5	KMC: B subC: B1 subC: B2 subC: B6 subC: B7	KMC: C subC: C1 subC: C2	KMC: D subC: D4 subC: D5

SUB OQE has created a set of 18 QTs: with term A generating equivalent class E and subclass A1-A5; B generated four subclasses B1, B2, B6 and B7; C and D harvested two subclasses respectively.

(ii) SUPER OQE, extends on the basis of (i) and add all "direct" super class of each KMC, i.e. it includes all direct super classes but not their sub branches, e.g. in Fig 17, direct superclasses of class B are B3 and B5 only, B4 and C3 as their subclasses are ignored. The traversal also assumes that the top-most superclass is a named class, i.e.

not the root node "Thing". Using SUPER OQE mode, the ontology traversal outcomes are shown in Table 2.

Table 2 SUB_SUPER OQE traversal outcomes

KMC	A	B	C	D
OQE Results	KMC: A EquC: E superC: F subC: A1 subC: A2 subC: A3 subC: A4 subC: A5	KMC: B superC: B3 superC: B5 subC: B1 subC: B2 subC: B6 subC: B7	KMC: C superC: C3 superC: B5 subC: C1 subC: C2	KMC: D superC: D1 superC: D2 superC: D3 subC: D4 subC: D5

The SUPER OQE on the basis of SUB OQE added 8 more super classes to the QT sets. The ontology traversal generated a QT set of 26 terms. It should be pointed out that B5, as the super class for both B and C, will be removed in a filtering stage that will result in 25 terms in QT set. The inclusion of duplicated filters after the OQE are necessary for limiting document search iterations and ensure that LMST's subsequent scoring algorithm will not duplicate term scores and inflate document relevance scores.

In order to clearly demonstrate the OQE process, it assumes that all the user's query inputs can be matched with ontology classes in the example. However, for the search experiments, provision will be made to accept input terms that may not feature in a specific ontology context but which may be related to the context of the target document.

Pseudo Code for Inheritance Class Hierarchy Algorithm

The inheritance class hierarchy algorithm generates query expansion terms depending on the selected class hierarchy expansion mode, i.e. SUB or SUPER. Fig 18 demonstrates the pseudo code for inheritance class hierarchy algorithm. The algorithm will also identify equivalent classes and individuals.

```
for each keyword k {
    add k to Query Array;
    for each ontology class c {
        if c equals k {
            if c has equivalent class c_eql {
                for c list c_eql  {add c_eql to Query Array.}
                for c list c_eql individuals{add individual to Query Array.}}
            else if subclass c_sub required AND c has c_sub {
                for c list c_sub {add c_sub to Query Array.}
                for c list subclass individuals{add individual to Query Array.}}
            else if superclass c^sup required AND c has c^sup AND NOT "Thing" AND NOT anonymous {
                for c list c^sup {add c c^sup to Query Array.
                    for c^sup list equivalent class sup^eql{add sup^eql to Query Array.}
                    for sup^eql list super-equivalent individuals{add individual to Query Array.}
                }
                for c list superclass individuals{add individual to Query Array.}}
        }
    }
}
```

Fig. 18 Pseudo Code for Inheritance Class Hierarchy Algorithm

The algorithm works in two stages: first, a user's input keyword is added to the query array, then it checks each base query term against the ontology classes. Then according to the selected options, OQE process adds corresponding equivalent classes, subclass, superclass and their respective class individual KMC to the query array.

4.3 Formulation of Concept Weights

Even for modular ontology based query expansion, the expanded term might be an over generalisation. It might be caused by two aspects: firstly, ambiguous terms might still exist within a specific ontology; secondly, even within a particular domain, users will differ in their understanding and information seeking behaviour.

The results of preliminary experiments stated this as the reason for assigning the term weighting to control the query expansion process. Different weighting functions produce different results, therefore there are two different term weighting approaches that collaborate with OQE respectively, i.e. entropy based term weighting and tf-idf term weighting.

(i) Entropy based term weighting

Entropy/log-entropy based term weighting is used to ensure the search fidelity, i.e. the expanded query term should not deviate too much from the original user's information requirement represented by the keyword. The expression of the word *w* is (Landauer et al., 2013):

$$weight_w = 1 - \varepsilon_w \qquad (10)$$

And

$$\varepsilon_w = \frac{1}{logN} \sum_{j=1}^{N} \frac{c(w, D_j)}{c(w, Coll)} \, log \frac{c(w, D_j)}{c(w, Coll)} \qquad (11)$$

Where

ε_w : normalized entropy of w occurs in collection;

N: total number of documents in the collection;

c(w,D_j): number of times w occurs in D_j;

c(w, Coll): number of time w occurs in the collection.

The entropy relies on two parts, the relative frequency of the term within the entire collection of documents *c(w, Coll)*, and relative frequency of a term in a document *c(w,Dj)*. By definition, $0 \leq \varepsilon_w \leq 1$, a value of ε_w close to 0 indicate that a word is distributed across only a few specific documents, while a value of ε_w close to 1 means that the word is presented in many documents. Therefore, the term weighing 1-ε_w indicates the indexing power of the word *w*.

(ii) Tf-idf based term weighting

Term Frequency (tf), is the parameter referring to the number of times a term is repeated in the document. Inverse document frequency (idf), is the parameter reflecting the "importance" of the term; the more documents that contain the term the less importance, so the value assigned to the term weighting is less. The algorithm can be represented in the following formula 12:

$$W_{tf\_idf} = \mathrm{tf} * \mathrm{idf} = \mathrm{tf} * \log \frac{\mathrm{N}}{\mathrm{df}} \tag{12}$$

Where

tf: term frequency in the document;

N: the total number of documents in the collection;

df: the number of documents that contain the term.

In the formula above, $\log \frac{\mathrm{N}}{\mathrm{df}}$ indicates idf, which adjusts the term weighting where it is common in the test collection.

4.4 Ontology based Document Classification

The ontology based document classification is based on the assumption that similar documents are relevant to the same query requests and should therefore represent a similar context. ODC utilises modular ontology concepts and hierarchy relationships, to classify the document used for local smoothing.

This subsection will discuss the ontology traversal executed in LMST's ODC process using the same example ontology used in section 4.2 (Fig 17). In contrast to the OQE example, it is assumed that part of the user's input keyword can be matched to the ontology concepts, i.e. four input keywords, A, B, D and K; the first three are KMCs and K is a generic term, i.e. not a class in the specific ontology.

The ODC process classifies the document collection in two or six different categories, i.e. it classifies the document collection into potentially relevant and non-relevant categories (**base** ODC, **sub** ODC and **super** ODC), or classifies the collection relying on the ontology class hierarchy structure (**separate** classify and **separate_FC** classify considering term frequency), based on the document classification mode option required.

(i) base ODC, implies classifying the documents which contain the initial keywords or the equivalent classes of KMC to a potentially relevant category. Otherwise, the document is classified as a non-relevant category. The flow chart in Fig 19 demonstrates the base ODC process.

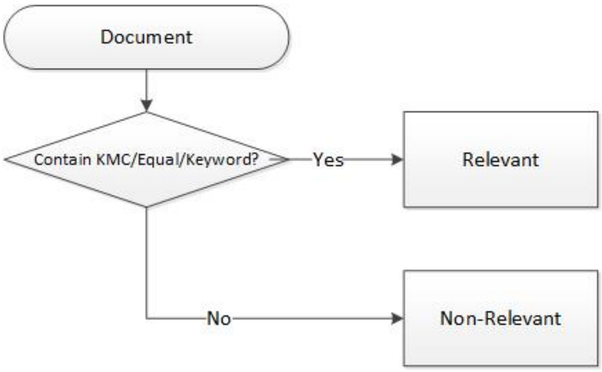


Fig. 19 Flow chart of Base ODC process

In Fig 20, the keyword and equivalent-class have been highlighted in green. It should be pointed out that the term ***k*** is not a class of the example ontology, but it is a keyword to determine the potentially relevant documents. All documents containing the term highlighted in green are classified into a relevant category.

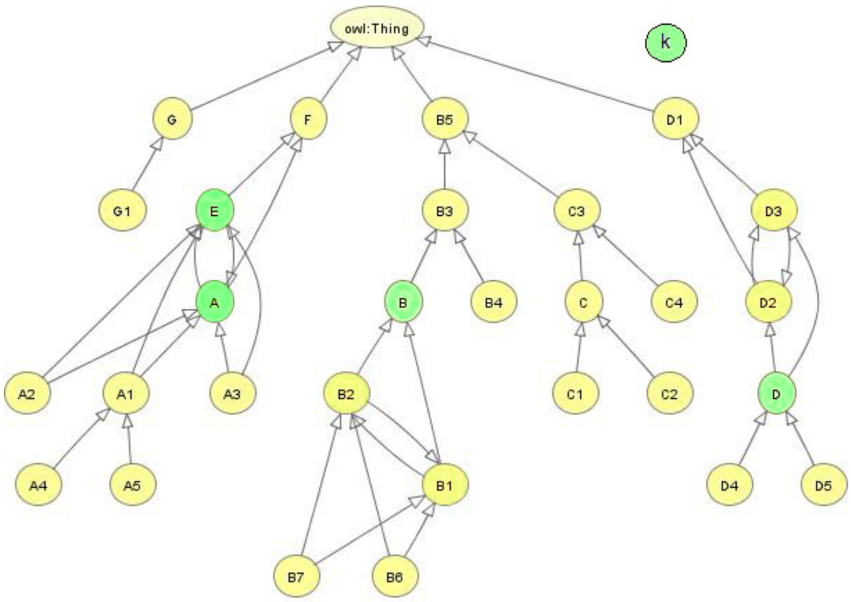


Fig. 20 Ontology example for Base ODC

(ii) sub ODC, implies on the basis of base ODC, that those documents which contain the subclass of KMC are classified as relevant documents. The Sub ODC process is represented in Fig 21.

The terms used to identify the document in the potentially relevant document category are denoted in green in Fig 22. The subclasses are added to the term sets compared with base ODC.

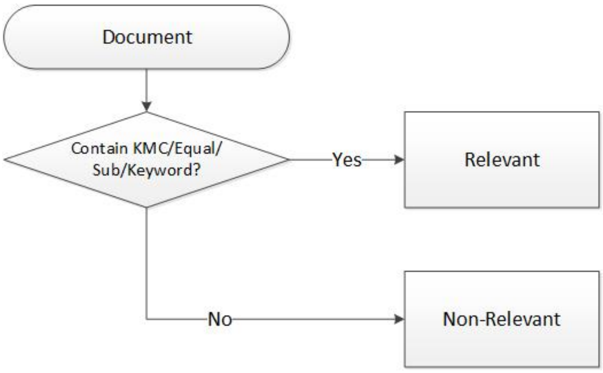


Fig. 21 Flow chart of Sub ODC process

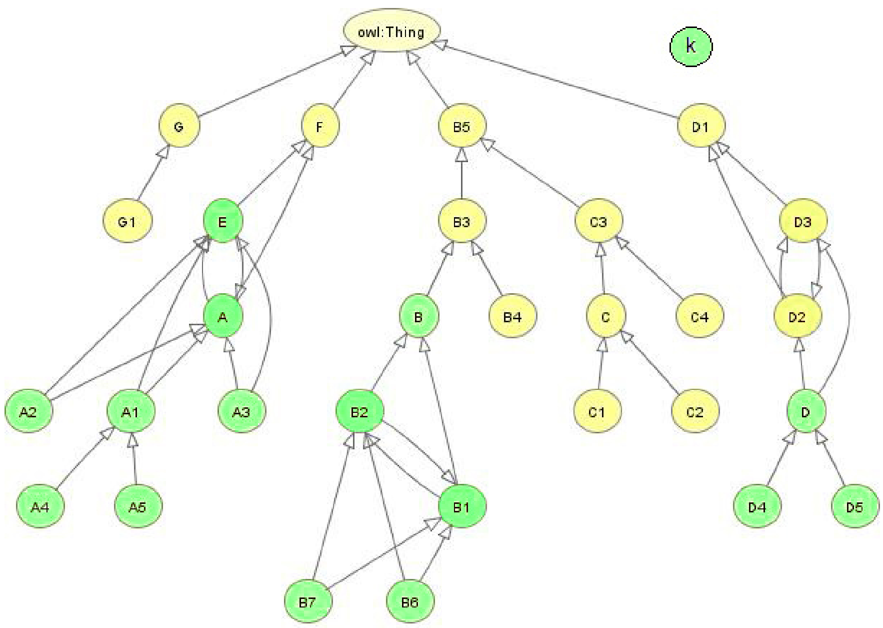


Fig. 22 Ontology example for Sub ODC

(iii) **super ODC,** implies that the document collection is divided into relevant category and non-relevant category. The document that contains the user's keyword and/or subclass of KMC and/or superclass of KMC is classified into the relevant category. Conversely, other documents are identified as non-relevant. On the basis of sub ODC, all the direct superclasses are added into the term sets. The super ODC process is illustrated in Fig 23 and the terms used to classify the relevant document are denoted in green in Fig 24.

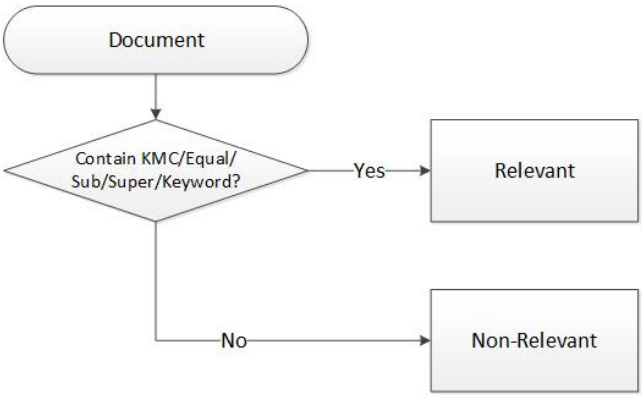


Fig. 23 Flow chart of Super ODC process

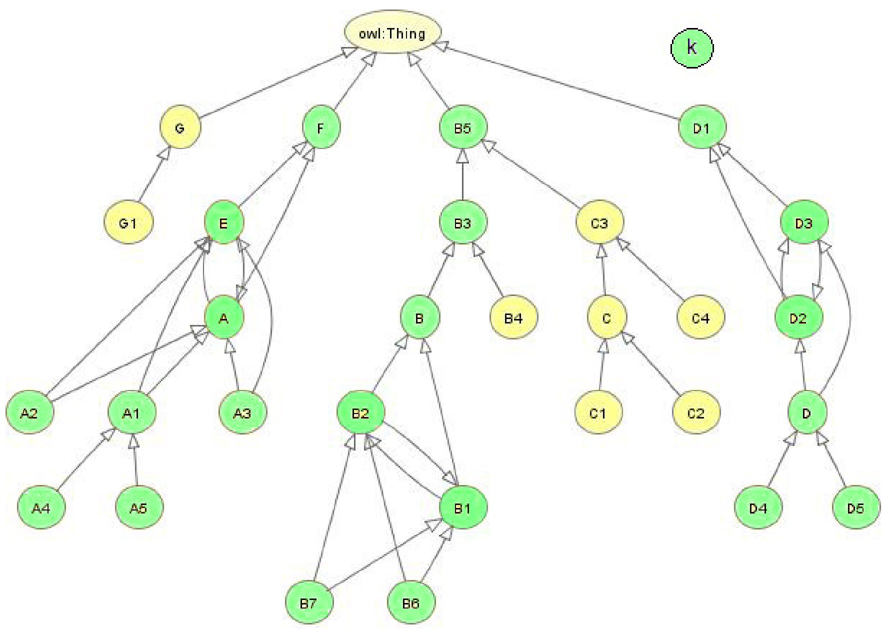


Fig. 24 Ontology example for Super ODC

(iv) ontology based separate classify (referred to as **separate**), divides the document collection into six categories. It determines the document categories based on whether or not the document contains keywords or different ontology hierarchy classes:

(1) Base category - all the documents containing the KMC, or the equivalent classes of KMC, are highlighted in green in Fig 25; (2) Sub category - includes those documents containing the subclass of KMC, highlighted in red; (3) Super category - documents containing the direct superclass of KMC are classified in this category - the term is denoted in blue; (4) Ontology classes category - all the documents containing any ontology classes, excluding the KMC, and their subclasses and superclass, are highlighted in yellow; (5) Keyword category - the documents only containing the keyword that is not matched with the ontology classes are classified to these categories – as denoted in white; (6) the remaining documents are in a potentially non-relevant category.

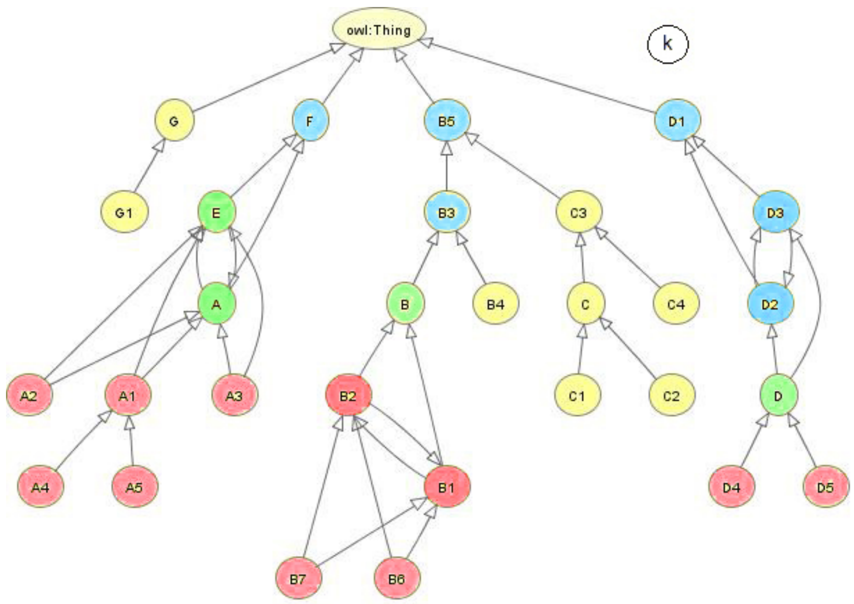


Fig. 25 Ontology example for Separate ODC

To eliminate the possibility of duplicated document calculations, the ODC works on the assumption that a document can only belong to one category. To correctly identify the categories of documents which contain the term in a different ontology hierarchy, a different priority be used to differentiate categories. The flow chart in Fig 26 shows the priority of the Separate ODC process: base > sub > super > ontology class > keyword > non-relevant. For instance, if the initial query is A, B, D and K, and it shows the document contains subclass A1, superclass B3 and keyword K, the document would map to *sub* category since A1 has the highest priority.

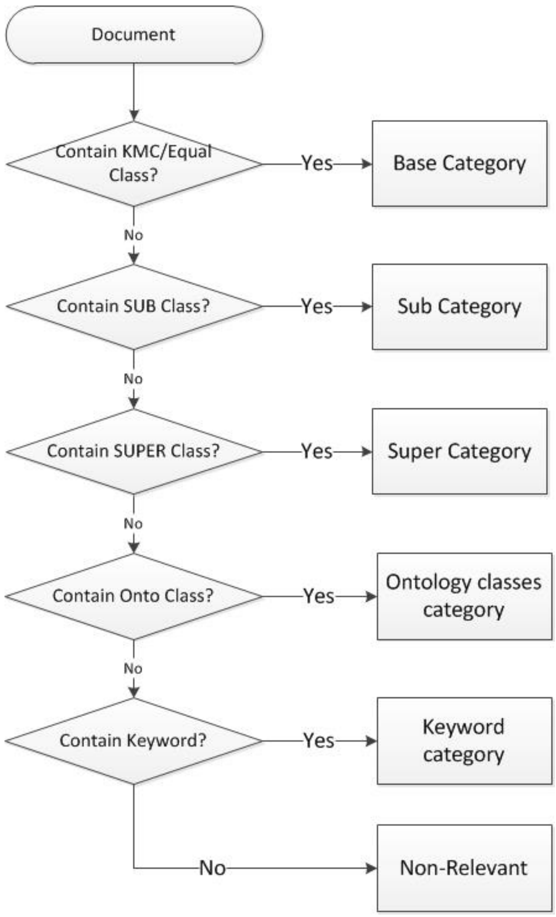


Fig. 26 Flow chart of Separate ODC process

(v) ontology based term frequency separate classify (separate_FC), - implies the basis of separate ODC, where the most occurring justify terms are utilised to classify the documents. The ontology traversal process is same with separate ODC (Fig 25), however the most occurring term is the key to mapping the document into this category. The separate_FC ODC process is illustrated in the flow chart in Fig 27.

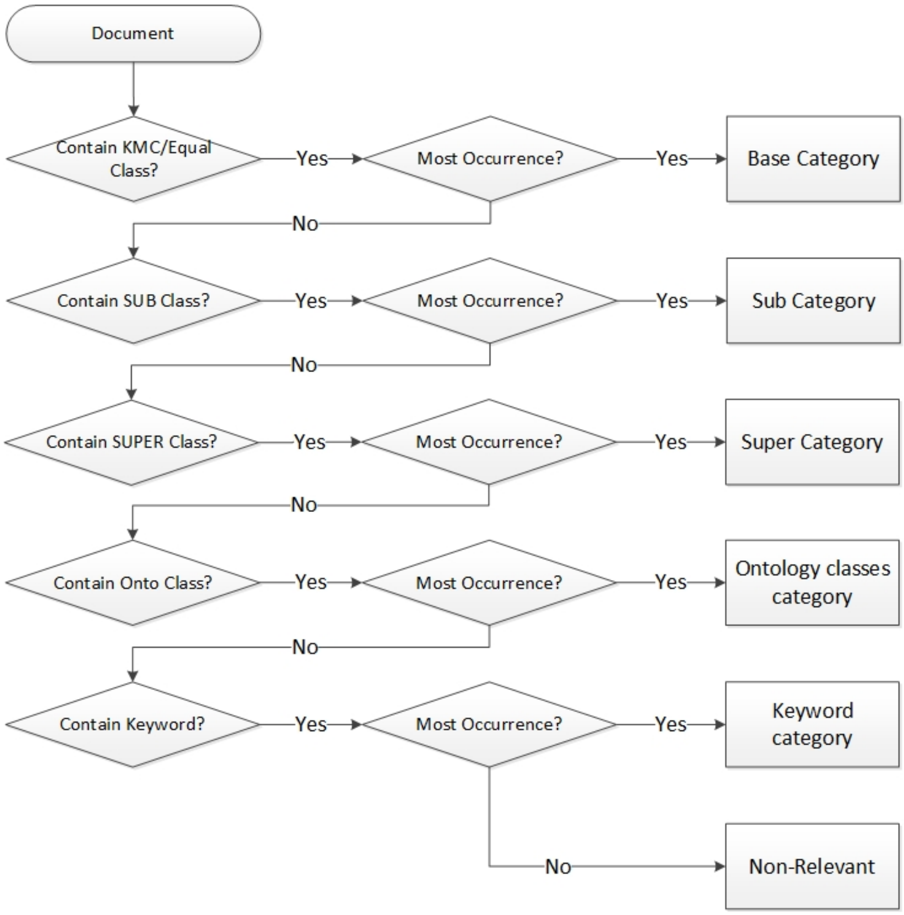


Fig. 27 Flow chart of Separate_FC ODC process

The above examples show how LMST's ODC algorithm would classify a document in terms of different ODC options. After the ODC process, the total number of times of query occurrence and category length (total number of the word in the category) will have been calculated for the smoothing and document scoring purpose (see section 4.5).

Pseudo Code for Class Hierarchy Document Classification Algorithm

The Ontology hierarchy based document classification algorithm classifies the document in the collection, depending on the selected ODC mode. The algorithm can be explained briefly: the algorithm first generates the keyword according to the user's input; the process is similar to the OQE process, corresponding ontology classes are added to the category term array. In terms of user's ODC option, it classifies the document into six categories, i.e. base, sub, super, onto, key and non-relevant. During the document classify process, the cumulative number of the query term occurrence (KO or OQE results) of each category would be counted, and total word count for every category will be calculated for scoring purposes.

```
for each keyword k {
    for each ontology class c {
        if c equals k {add c to keyword array.
            if c has c_eql {
                for c list c_eql {add c_eql to keyword array.}
                for c list c_eql individuals {add individual to keyword array.}}
            else if has c_sub {
                for c list c_sub {add c_sub to keyword array.}
                for c list c_sub individuals {add individual to keyword array.}}
        }
        else {add k to keyword array.}
    } }
for each document d {
    if Base-classify or Sub-classify or Onto-classify is required {
        for each keyword in keyword array {
            if d contains c or c-individual or c_eql or c_eql-individual {
                document classified to base-category Cate_base.
                add document length to Cate_base length.
                add query count to Cate_base total quay cont.}
            else if d contains c_sub or c_sub-individual {
                document classified to sub- category Cate_sub.
                add document lenght to Cate_sub length.
                add query count to Cate_sub total quay cont.}
            else if d contains ontology class {
                document classified to onto- category Cate_onto.
                add document lenght to Cate_onto length.
                add query count to Cate_onto total quay cont.}
            else if d contains k {
                document classified to key- category Cate_key.
                add document lenght to Cate_key length.
                add query count to Cate_key total quay cont.}
        } }
    else {add document lenght to collection length.
        add query count to collection total quay cont.  }
}
```

Fig. 28 Pseudo Code for Class Hierarchy Document Classification Algorithm

4.5 Ontology enhanced language model

The main task of a text ad hoc IR system is to capture relevant documents according to a user's input query, essentially rank the documents based on document score, so that highly relevant documents are listed above the less relevant documents and non-relevant ones. Clearly, the performance of an IR system is primarily determined by the soundness of its underlying retrieval model, which defines the notion of relevance and enables the IR system to compute the score to rank documents.

The basic approach for using SLM for IR is to model the query generation, estimate a language model for each document, and rank documents according to likelihood the query Q could have been generated from each of these document models, i.e. $P(Q|D)$. $P(Q|D)$ can be represented as a product of the individual query term probability, i.e.

$$P(Q|D) = \prod_{i=1}^{m} P(w_i|D) \qquad (13)$$

Where w_i is the i^{th} word in the query and $P(w|D)$ is calculated by the smoothed document language model (Zhai, 2009):

$$P(w|D) = \lambda P_{ML}(w|D) + (1-\lambda) P_{ML}(w|Coll) \qquad (14)$$

Where, $P_{ML}(w|D)$ indicates maximum likelihood estimate of word probability of word w in document D;

$P_{ML}(w|Coll)$ indicates maximum likelihood estimate of word probability of word w in the entire collection *Coll*;

λ is a coefficient controlling sum of two probabilities less than 1, and generally depends on the document or an arbitrary weight between 0 and 1. It takes different forms when different smoothing methods are used.

Primarily research state that Bayesian Dirichlet prior (or short for Dirichlet) smoothing produce the best interpolation smoothing results for short query search (Zhai and

Lafferty, 2004). The λ takes the form:

$$\lambda = \frac{|D|}{|D| + \mu} \tag{15}$$

Where, μ is the Dirichlet smoothing parameter.

Dirichlet smoothing estimates the probability of an unseen word, based on its entire collection probability. Formula (14) represents the pure language model with Dirichlet smoothing and is used as the baseline of the search performance comparison experiments. The objective of this research is to develop new search models for IR in the language model framework and evaluate the performance of these models using P&R measure. Enhanced search models will be achieved from two approaches, i.e. OQE and ODC. The OQE process provides higher term probability to the potential document which contains the extension terms, and the ODC process provides a more accurate local probability estimation using category structure. Meanwhile, entropy or tf-idf is exploited to control the quality of the OQE.

Fig 29 shows the combinations of OQE options, term weighting options and ODC options. The following subsection will demonstrate these established search modes in detail.

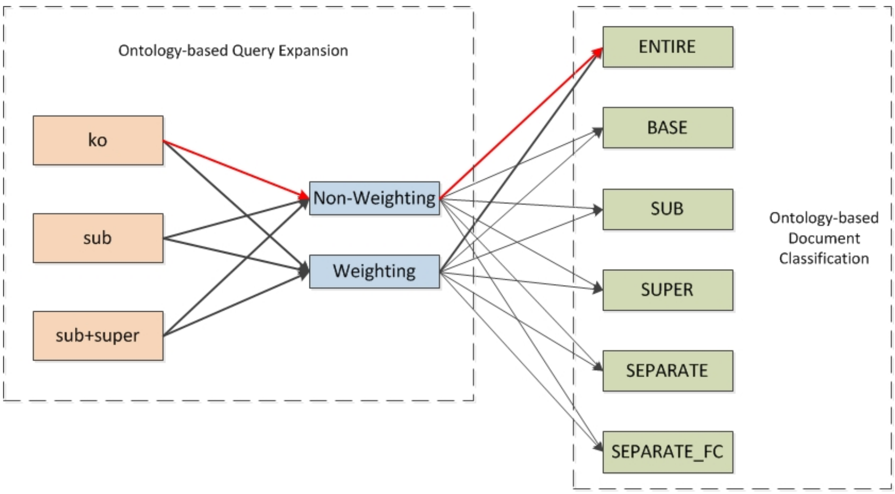


Fig. 29 Combinations of OQE, Weighting and ODC

It should be pointed out that the search mode is named according to the corresponding OQE (uppercase letter) process and ODC (lowercase letter) processes, connected with

"+", e.g. baseline represented by formula (14), denoted in red in Fig 29, can be presented as KO+entire. It indicates that the system searches for the keyword-only and then smoothing the unseen word according to its probability distribution in the entire collection. Since the entropy weighting and tf-idf weighting have same scoring formula, for clarity, the two weighting approaches are merged represented as *weighting* option.

4.5.1 OQE-only enhanced search modes

In this search mode set, OQE is the only approach exploited to enhance the pure language model so that non-weighing and the smoothing processes are the same as the baseline. It takes a similar approach for formula (14) by combining the probability of KMC and its corresponding ontology expanded terms.

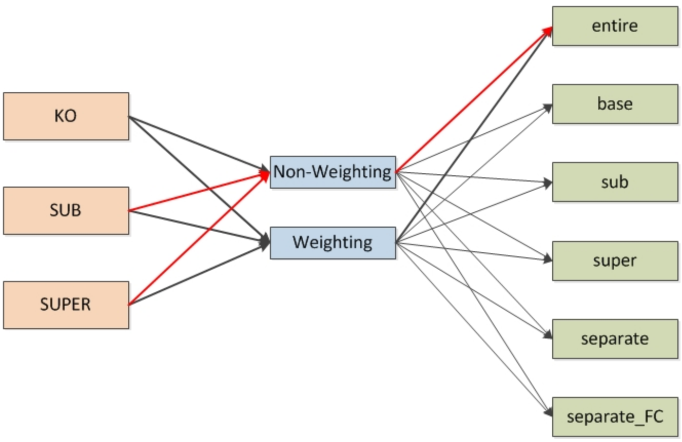


Fig. 30 OQE enhanced search model

Fig 30 highlights the combinations of three approaches in red, i.e. SUB+entire and SUPER+entire.

SUB+entire search mode

Similar to the baseline, the SUB+entire calculates p(w|D) based on the probability of a keyword and its corresponding subclass in the document and collection, i.e. $P_{ML}(w \cup w_{sub}|D)$ and $P_{ML}(w \cup w_{sub}|Coll)$ as shown in formula (16).

$$\begin{aligned} \mathrm{P(w,D)} &= \lambda P(w \cup w_{sub}|D) + (1-\lambda)P(w \cup w_{sub}|Coll) \\ &= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda)\frac{c(w \cup w_{sub}, Coll)}{|Coll|} \end{aligned} \tag{16}$$

Where, $c(w \cup w_{sub},D)$ indicates the sum number of times term w and its subclass w_{sub} occurs in the document D;

$c(w \cup w_{sub},Coll)$ indicates the sum number of times term w and its subclass w_{sub} occurs in the collection *Coll*;

$|D|$ indicates the total number of words in D;

$|Coll|$ indicates the total number of words in *Coll*.

SUPER + entire search mode

Compared with SUB+entire, SUPER+entire adds corresponding superclasses to the query term set and uses it to calculate the probability of query in the documents. The formula is shown below:

$$\begin{aligned} \mathrm{P(w,D)} &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)P(w \cup w_{sub} \cup w^{sup}|Coll) \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\frac{c(w \cup w_{sub} \cup w^{sup}, Coll)}{|Coll|} \end{aligned} \tag{17}$$

4.5.2 ODC-only enhanced search modes

In this set of search modes, the system searches for keyword-only against the document corpus. However, instead of the smoothing language model using the corresponding probabilistic distribution of unseen word in the entire collection, the category feature is exploited in the smoothing process, i.e. probabilities of unseen words in the category are used to determine the $P(w|D)$. The categories are assumed to be relevant to the topic, which are described in different hierarchies of the ontology classes. The combinations between keyword-only and five different ODC approaches are denoted in red in Fig 31.

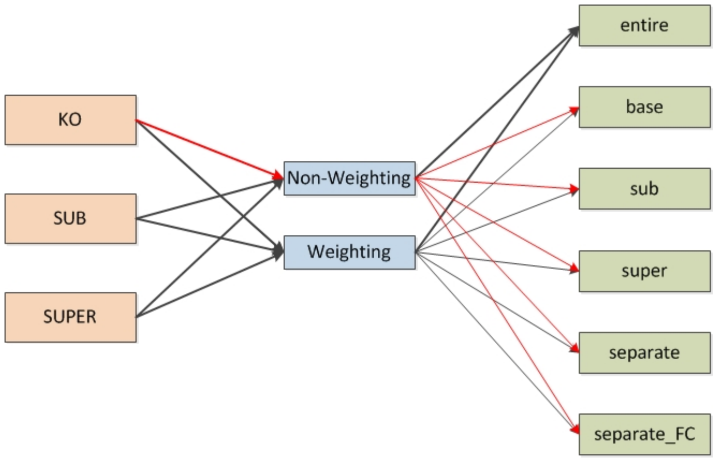


Fig. 31 ODC enhanced search model

KO+base search mode

KO+base exploits the category feature for smoothing, which is classified by keyword and its corresponding equivalent classes in the ontology context, to estimate the probability of the unseen word.

$$P(w|D) = \lambda P_{ML}(w|D) + (1-\lambda)P_{ML}(w|Cate_{Base}) = \lambda\frac{c(w,D)}{|D|} + (1-\lambda)\frac{c(w,Cate_{Base})}{|Cate_{Base}|} \quad (18)$$

Where, $c(w,Cate_{Base})$ is the number of times term w occurs in the *Base* category,

$|Cate_{base}|$ is the total number of words in *Base* category $Cate_{base}$.

KO+sub search mode

This approach exploits Sub category for language model smoothing.

$$P(w|D) = \lambda P_{ML}(w|D) + (1-\lambda)P_{ML}(w|Cate_{Sub}) = \lambda\frac{c(w,D)}{|D|} + (1-\lambda)\frac{c(w,Cate_{Sub})}{|Cate_{Sub}|} \quad (19)$$

Where, $c(w,Cate_{sub})$ is the number of times term w occurs in the *Sub* category,

$|Cate_{sub}|$ is the total number of words in *Sub* category $Cate_{sub}$.

KO+super search mode

In this search mode, superclasses provide more potential documents to estimate the probability of unseen terms in the documents:

$$\mathrm{P(w,D)} = \lambda P(w|D) + (1-\lambda)P\left(w\middle|Cate_{Super}\right) = \lambda\frac{c(w,D)}{|D|} + (1-\lambda)\frac{c(w,Cate_{Super})}{|Cate_{Super}|} \quad (20)$$

Where, $c(w,Cate_{Super})$ is the number of times term w occurs in the *Super* category,

$|Cate_{super}|$ is the total number of words in *Base* category $Cate_{super}$.

KO+separate search mode

In contrast to previous ODC approaches which classify the documents into relevant or non-relevant categories. The separate ODC approach divides the document collation into six categories. The separate ODC enhanced smoothing algorithm estimates the document model in two stages. The first stage estimates the category language model based on the entire collection, where the basic idea is to combine the documents in the same category and treat the category as if it were a long document. Category language model can be estimated using the formula (21):

$$P(w|Cate) = \beta P_{ML}(w|Cate) + (1-\beta)P_{ML}(w|Coll) \quad (21)$$

Where β is the coefficient controlled sum of two probabilities less than 1. It models the probability of word w in the category. The second stage estimates the document language model based on the category language model estimated result from stage one. The algorithm can be represented as:

$$\begin{aligned} P(w|D) &= \lambda P_{ML}(w|D) + (1-\lambda)[\beta P_{ML}(w|Cate_D) + (1-\beta)P_{ML}(w|Coll)] \\ &= \lambda\frac{c(w,D)}{|D|} + (1-\lambda)\left[\beta\frac{c(w,Cate_D)}{|Cate_D|} + (1-\beta)\frac{c(w,Coll)}{|Coll|}\right] \end{aligned} \quad (22)$$

Where, $c(w,Cate_D)$ is the number of times term w occurs in the document D located category $Cate_D$,

$|Cate_D|$ is the total number of words in $Cate_D$.

β is smoothing coefficient of category smoothing, in this search, the value of β utilise the constant 0.7, since it provide best search results in Liu and Croft's experiments (Liu and Croft, 2004).

KO+separate_FC search mode

Compared with KO+separate, KO+separate_FC has a different document classification approach, i.e. to classify the document according to the most occurrence query term. It differs from KO+separate, as the document category in this model is represented as *Cate'*.

$$\begin{aligned}\mathrm{P(w,D)} &= \lambda P(w|D) + (1-\lambda)[\beta P(w|Cate'_D) + (1-\beta)P(w|Coll)] \\ &= \lambda \frac{c(w,D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w,Cate'_D)}{|Cate'_D|} + (1-\beta)\frac{c(w,Coll)}{|Coll|}\right]\end{aligned} \quad (23)$$

Where, $c(w,Cate'_D)$ is the number of times term *w* occurs in the most occurrence term located category $Cate'_D$,

$|Cate'_D|$ is the total number of words in $Cate'_D$.

4.5.3 Weighting enhanced search mode

Term weighing is exploited to solve the problem that a query term might be a generalisation term. The objective of term weighing is allocate a higher weighing to more specific query terms and a lower weight to a general one - to control the quality of query terms.

KO + entire with term weighting

This is different to the pure language model represented in formula (14), the term weighing e_w is applied to reduce the real term count to some extent in the document:

$$
\begin{aligned}
\mathrm{P(w, D)} &= \lambda P(e_w * w|D) + (1-\lambda) P(e_w * w|Coll) \\
&= \lambda \frac{e_w * c(w, D)}{|D|} + (1-\lambda) \frac{e_w * c(w, Coll)}{|Coll|}
\end{aligned} \tag{24}
$$

Where, e_w is the term weighing of term *w*.

4.5.4 OQE + ODC search modes

As all the OQE-only, ODC-only and weighting only enhanced language models have been discussed in the previous subsection, the formulas for the remaining 27 combinations of OQE, ODC and weighing search modes are shown in this subsection.

SUB + base

$$
\begin{aligned}
\mathrm{P(W, D)} &= \lambda P(w \cup w_{sub} |D) + (1-\lambda) P(w \cup w_{sub} |Cate_{Base}) \\
&= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda) \frac{c(w \cup w_{sub}, Cate_{Base})}{|Cate_{Base}|}
\end{aligned} \tag{25}
$$

SUB + sub

$$
\begin{aligned}
\mathrm{P(W, D)} &= \lambda P(w \cup w_{sub} |D) + (1-\lambda) P(w \cup w_{sub} |Cate_{Sub}) \\
&= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda) \frac{c(w \cup w_{sub}, Cate_{Sub})}{|Cate_{Sub}|}
\end{aligned} \tag{26}
$$

SUB + super

$$
\begin{aligned}
\mathrm{P(W, D)} &= \lambda P(w \cup w_{sub} |D) + (1-\lambda) P\left(w \cup w_{sub} \middle| Cate_{Super}\right) \\
&= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda) \frac{c(w \cup w_{sub}, Cate_{Super})}{|Cate_{Super}|}
\end{aligned} \tag{27}
$$

SUB + separate

$$
\begin{aligned}
\mathrm{P(W, D)} &= \lambda P(w \cup w_{sub}|D) + (1-\lambda)[\beta P(w \cup w_{sub}|Cate_D) + (1-\beta) P(w \cup w_{sub}|Coll)] \\
&= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda) \left[\beta \frac{c(w \cup w_{sub}, Cate_D)}{|Cate_D|} + (1-\beta) \frac{c(w \cup w_{sub}, Coll)}{|Coll|}\right]
\end{aligned} \tag{28}
$$

SUB + separate_FC

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(w \cup w_{sub}|D) + (1-\lambda)[\beta P(w \cup w_{sub}|Cate'_{D}) + (1-\beta)P(w \cup w_{sub}|Coll)] \\ &= \lambda \frac{c(w \cup w_{sub}, D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w \cup w_{sub}, Cate'_D)}{|Cate'_D|} + (1-\beta)\frac{c(w \cup w_{sub}, Coll)}{|Coll|}\right]\end{aligned} \tag{29}$$

SUPER + base

$$\begin{aligned}P(w,D) &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)P(w \cup w_{sub} \cup w^{sup}|Cate_{Base}) \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\frac{c(w \cup w_{sub} \cup w^{sup}, Cate_{Base})}{|Cate_{Base}|}\end{aligned} \tag{30}$$

SUPER + sub

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)P(w \cup w_{sub} \cup w^{sup}|Cate_{Sub}) \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\frac{c(w \cup w_{sub} \cup w^{sup}, Cate_{Sub})}{|Cate_{Sub}|}\end{aligned} \tag{31}$$

SUPER + super

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)P\left(w \cup w_{sub} \cup w^{sup}\middle|Cate_{Super}\right) \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\frac{c(w \cup w_{sub} \cup w^{sup}, Cate_{Super})}{|Cate_{Super}|}\end{aligned} \tag{32}$$

SUPER + separate

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)[\beta P(w \cup w_{sub} \cup w^{sup}|Cate_D) + (1-\beta)P(w \cup w_{sub} \cup w^{sup}|Coll)] \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w \cup w_{sub} \cup w^{sup}, Cate_D)}{|Cate_D|} + (1-\beta)\frac{c(w \cup w_{sub} \cup w^{sup}, Coll)}{|Coll|}\right]\end{aligned} \tag{33}$$

SUPER + separate_FC

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(w \cup w_{sub} \cup w^{sup}|D) + (1-\lambda)[\beta P(w \cup w_{sub} \cup w^{sup}|Cate'_D) + (1-\beta)P(w \cup w_{sub} \cup w^{sup}|Coll)] \\ &= \lambda \frac{c(w \cup w_{sub} \cup w^{sup}, D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w \cup w_{sub} \cup w^{sup}, Cate'_D)}{|Cate'_D|} + (1-\beta)\frac{c(w \cup w_{sub} \cup w^{sup}, Coll)}{|Coll|}\right]\end{aligned} \tag{34}$$

KO + base with weight

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(e_w * w|D) + (1-\lambda)P(e_w * w|Cate_{Base}) \\ &= \lambda \frac{e_w * c(w,D)}{|D|} + (1-\lambda)\frac{e_w * c(w,Cate_{Base})}{|Cate_{Base}|}\end{aligned} \quad (35)$$

KO + sub with weight

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(e_w * w|D) + (1-\lambda)P(e_w * w|Cate_{Sub}) \\ &= \lambda \frac{e_w * c(w,D)}{|D|} + (1-\lambda)\frac{e_w * c(w,Cate_{Sub})}{|Cate_{Sub}|}\end{aligned} \quad (36)$$

KO + super with weight

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(e_w * w|D) + (1-\lambda)P\left(e_w * w\middle|Cate_{Super}\right) \\ &= \lambda \frac{e_w * c(w,D)}{|D|} + (1-\lambda)\frac{e_w * c(w,Cate_{Super})}{|Cate_{Super}|}\end{aligned} \quad (37)$$

KO + separate with weight

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(e_w * w|D) + (1-\lambda)[\beta P(e_w * w|Cate_D) + (1-\beta)P(e_w * w|Coll)] \\ &= \lambda \frac{e_w * c(w,D)}{|D|} + (1-\lambda)\left[\beta \frac{e_w * c(w,Cate_D)}{|Cate_D|} + (1-\beta)\frac{e_w * c(w,Coll)}{|Coll|}\right]\end{aligned} \quad (38)$$

KO + separate_FC with weight

$$\begin{aligned}\mathrm{P(W,D)} &= \lambda P(e_w * w|D) + (1-\lambda)[\beta P(e_w * w|Cate'_D) + (1-\beta)P(e_w * w|Coll)] \\ &= \lambda \frac{e_w * c(w,D)}{|D|} + (1-\lambda)\left[\beta \frac{e_w * c(w,Cate'_D)}{|Cate_D|} + (1-\beta)\frac{e_w * c(w,Coll)}{|Coll|}\right]\end{aligned} \quad (39)$$

SUB + entire with weight

$$\mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub}|D) + (1-\lambda)P(w \cup e_{sub} * w_{sub}|Coll) \quad (40)$$

SUB + base with weight

$$\mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub}|D) + (1-\lambda)P(w \cup e_{sub} * w_{sub}|Cate_{Base}) \quad (41)$$

SUB + sub with weight

$$\mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \,|D) + (1-\lambda) P(w \cup e_{sub} * w_{sub} \,|Cate_{Sub}) \tag{42}$$

SUB + super with weight

$$\mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \,|D) + (1-\lambda) P\big(w \cup e_{sub} * w_{sub} \,\big|Cate_{Super}\big) \tag{43}$$

SUB + separate with weight

$$\begin{aligned} \mathrm{P(W,D)} &= \lambda P(w \cup e_{sub} * w_{sub} |D) + (1-\lambda)[\beta P(w \cup e_{sub} * w_{sub} |Cate_D) + (1-\beta) P(w \cup e_{sub} * w_{sub} |Coll)] \\ &= \lambda \frac{c(w \cup e_{sub} * w_{sub}, D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w \cup e_{sub} * w_{sub}, Cate_D)}{|Cate_D|} + (1-\beta)\frac{c(w \cup e_{sub} * w_{sub}, Coll)}{|Coll|}\right] \end{aligned} \tag{44}$$

SUB + separate_FC with weight

$$\begin{aligned} \mathrm{P(W,D)} &= \lambda P(w \cup e_{sub} * w_{sub} |D) + (1-\lambda)[\beta P(w \cup e_{sub} * w_{sub} |Cate'_D) + (1-\beta) P(w \cup e_{sub} * w_{sub} |Coll)] \\ &= \lambda \frac{c(w \cup e_{sub} * w_{sub}, D)}{|D|} + (1-\lambda)\left[\beta \frac{c(w \cup e_{sub} * w_{sub}, Cate'_D)}{|Cate'_D|} + (1-\beta)\frac{c(w \cup e_{sub} * w_{sub}, Coll)}{|Coll|}\right] \end{aligned} \tag{45}$$

SUPER + entire with weight

$$\mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|D) + (1-\lambda) P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|Coll) \tag{46}$$

SUPER + base with weight

$$\begin{aligned} \mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|D) \\ + (1-\lambda) P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|Cate_{Base}) \end{aligned} \tag{47}$$

SUPER + sub with weight

$$\begin{aligned} \mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|D) \\ + (1-\lambda) P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|Cate_{Sub}) \end{aligned} \tag{48}$$

SUPER + super with weight

$$\begin{aligned} \mathrm{P(W,D)} = \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,|D) \\ + (1-\lambda) P\big(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \,\big|Cate_{Super}\big) \end{aligned} \tag{49}$$

SUPER + separate with weight

$$\begin{aligned}\mathrm{P(W,D)} = {} & \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} | D) \\ & + (1-\lambda)[\beta P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} | Cate_D) \\ & + (1-\beta) P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} | Coll)]\end{aligned} \tag{50}$$

SUPER + separate_FC with weight

$$\begin{aligned}\mathrm{P(W,D)} = {} & \lambda P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} | D) \\ & + (1-\lambda)\left[\beta P\left(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} \middle| Cate'_{D}\right)\right. \\ & \left. + (1-\beta) P(w \cup e_{sub} * w_{sub} \cup e^{sup} * w^{sup} | Coll)\right]\end{aligned} \tag{51}$$

5. EXPERIMENTATION

5.1 Search effectiveness experiment steps

This chapter will outline the basic steps and assumptions considered for the proposed experiment. LMST returns documents which contain either the keywords alone or a context-driven expanded keyword expansion according to user's input. The document corpus will be pattern-matched against the search terms and search models used to calculate the document relevance ranking. As mentioned in chapter 3, the assessment of search performance would be focused primarily on precision outcomes in the 10% to 30% recall range.

The objective was to examine the impact of ontology based query expansion (OQE) and ontology based document classification (ODC) compared to keyword only SLM (see hypothesis, section 2.5). However, the design and construction of modular ontology is labour intensive and time consuming, therefore the experiments will be based on 5 TREC topics. In each experiment, 10 query sets related to the topic would be executed over whole document corpus, using different search modes. The selection of the topic and search query would be based on TREC query topic statements.

Schematic of the LMST process

The initialisation of the OQE, ODC and relevance measurement process is reflected in five key stages shown in Fig 32. It involves search mode selection and keyword entry (A), query term generation processing (B), smoothing (C), loading corpus and document text analysis using pattern matching (D), and relevance based document scoring algorithm for P&R (E). The processes differ in the query term generation stage and the document classification stage:

The OQE stage in (B) provides 3 options:

1. **Keyword-only** mode, which the user’s input terms are forwarded as the query term set.

2. Generate the ontology query expansion set, which include three modes, i.e. **SUB** OQE, **SUPER** OQE.

The weighting processes in (B) provides 2 options:

1. non-weighting, where the default weighing of each term in the query set is 1.

2. weighting, which includes two approaches, i.e. **entropy** and **tf-idf**.

The ODC in stage (C) has 6 options:

1. Dirichlet smoothing (**entire**), where the probability distribution of the unseen word is estimated using the entire document corpus.

2. ODC based smoothing, which includes five modes, i.e. **base** ODC, **sub** ODC, **super** ODC, **separate** ODC, and ontology based separate ODC with term frequency (**separate_FC**).

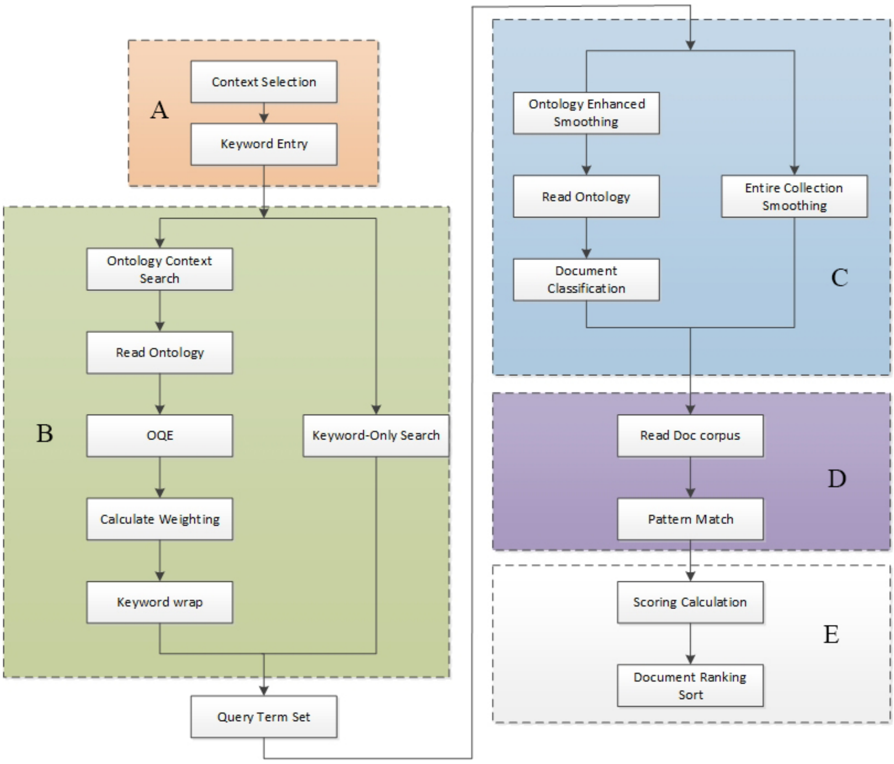


Fig. 32 Key LMST search, measurement and comparison process stages

The planned 54 search modes are produced by the user's search mode selection. Combinations from Stage B and Stage C are shown in Table 3.

Table 3 Combinations of keyword processes and smoothing approaches

		entire	base ODC	sub ODC	super ODC	separate ODC	separate_FC ODC
Non-weight	Keyword-Only	KO+entire	KO+base	KO+sub	KO+super	KO+separate	KO+separate_FC
	SUB OQE	SUB+entire	SUB+base	SUB+sub	SUB+super	SUB+separate	SUB+separate_FC
	SUPER OQE	SUPER+entire	SUPER+base	SUPER+sub	SUPER+super	SUPER+separate	SUPER+separate_FC
entropy	Keyword-Only	KO_e+entire	KO_e+base	KO_e+sub	KO_e+super	KO_e+separate	KO_e+separate_FC
	SUB OQE	SUB_e+entire	SUB_e+base	SUB_e+sub	SUB_e+super	SUB_e+separate	SUB_e+separate_FC
	SUPER OQE	SUPER_e+entire	SUPER_e+base	SUPER_e+sub	SUPER_e+super	SUPER_e+separate	SUPER_e+separate_FC
idf	Keyword-Only	KO_idf+entire	KO_idf+base	KO_idf+sub	KO_idf+super	KO_idf+separate	KO_idf+separate_FC
	SUB OQE	SUB_idf+entire	SUB_idf+base	SUB_idf+sub	SUB_idf+super	SUB_idf+separate	SUB_idf+separate_FC
	SUPER OQE	SUPER_idf+entire	SUPER_idf+base	SUPER_idf+sub	SUPER_idf+super	SUPER_idf+separate	SUPER_idf+separate_FC

Five experiment stages are discussed in the following five subsections.

5.1.1 Search Mode Selection and Keyword Entry (Stage A)

For the search effectiveness comparison experiments, it will be assumed that a number of context (topic-related) ontologies would be available to guide the user in semantic search, e.g. when a user starts to enter a base query term, the system would traverse a selected ontology to return concepts (i.e. "suggested query terms") that match the leading characters of input term – see Fig 37 (subsection 5.1.7). To achieve this, LMST would require a user to first select a context and then selected the classes from generated class hierarchy.

For the purpose of realistically and objectively estimating a user's search input, provision should be made for users to use generic/independent terms rather than the more specific "recommend text", i.e. allow input of terms that may not feature in a specific ontology context but which may be related to the context of the target document. For instance, to search the target document relevant to T401 (Foreign Minorities, Germany), the input term should contain the proper noun "Germany" to restrict the search range. However, "Germany" was not specified in the immigration ontology context. Further, the query narrative in the T416 "Three Gorges Project" (Fig 45 in subsection 5.2.3) seeks relevant documents containing "date of completion" information, but this generic term/phrase was not considered solely relevant to a Hydro-electric ontology. As shown in the section 5.2, a query term matrix can contain several generic terms that may not be solely relevant to one specific domain, i.e. not in any of the manually built context ontologies.

5.1.2 Query Term Generation Processing (Stage B)

In order to realistically simulate Web search, the input query terms would comprise a set of 4 base keywords or short phrases. The matrices of query sets, created by various combining topic relevant query terms, are shown in section 5.2.

Before a document is pattern matched against the search terms, in this stage, LMST will initialise the input query terms. It includes keyword-only and an OQE based query expansion process for disambiguation. Term weighing is also applied to eliminate content bias cause by query expansion.

Keyword-only (KO)

The keyword-only search is straightforward, i.e. the user's input terms are passed as the query term sets and the document corpus is systematically scanned for pattern-matches within each text repository.

Ontology-based query expansion (OQE)

It was envisaged that the process for OQE-based search would require an intermediary stage before searching. The user should be able to further control ontology query expansion by selecting options to determine the range of an expansion, i.e. users can choose equivalent class, subclass and superclass options for query expansion. According to the user's OQE selection, the specific ontology context would be traversed to find expanded query terms - see section 4.2.

Non-weighing

With this option, the weighting for every term in the query set is equal, i.e. 1. The counts for the expand terms are directly added to the initial query counts.

Weighting (entropy/ tf-idf)

The expanded query sets should not deviate too much from the original user's input keywords. Therefore, term weighting has been applied to control the query expansion and assure the process has "fidelity". For the purpose of evaluating search performance using different weightings, two weighting options are provided for a user to control the query expansion process, i.e. users can choose entropy and tf-idf options for term weighting processing.

5.1.3 Document Corpus Processing (Stage C)

In this stage, document corpus processing provides two strategies to estimate the document language model.

Entire document collection based global smoothing

The basic process for the entire document collection smoothing should be straightforward, which is typically by estimating the document language model based on probabilistic distribution of the entire document collection. The smoothing process replaces the zero probability of unseen terms according to its probability in the entire document collection.

Document classification based local smoothing

Ontology-based document classification leverages ontology structures to offer more accurate smoothing for each individual document. This is in contrast to the simple entire document collection smoothing strategy, where all documents are smoothed with the same collection model.

In this research, the user should able to further control the document classification by selecting the option to determine the range of ontology-based document classification, i.e. baseclass, subclass, superclass or whole ontology-based enhanced language model smoothing.

5.1.4 Search Term Pattern Matching and Validation (Stage D)

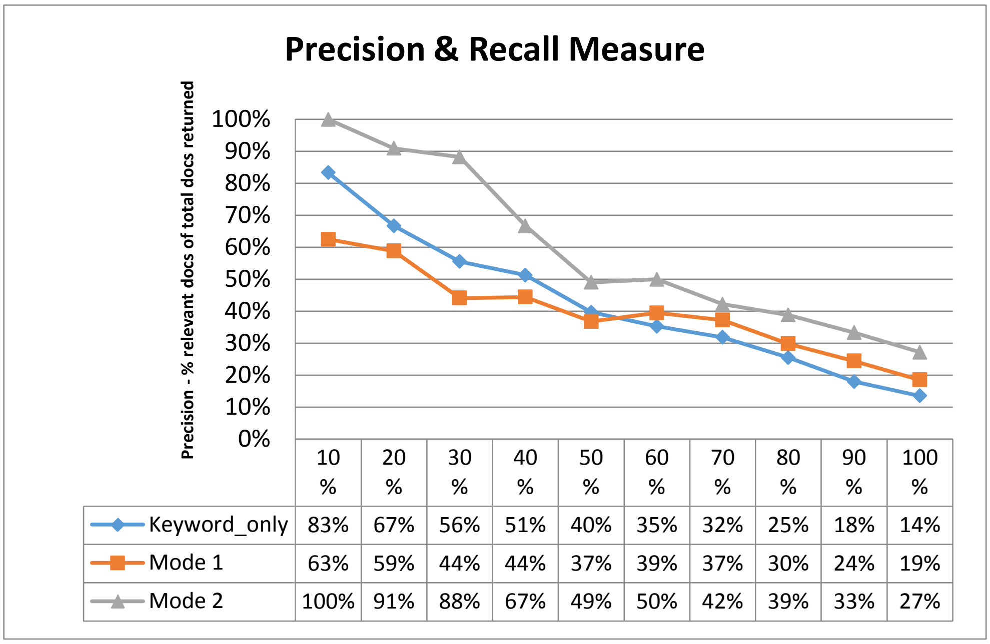
In stage D, generated query terms from stage C is pattern-matched to retrieve relevant documents which contain either the keywords alone or ontology-driven keyword expansion.

For the purpose of estimating the document model reliability, search trials were initially conducted using a small document set so that term frequency could be manually validated. Any matching issues were resolved by refining the regular expression syntax until terms were accurately identified and returned from the document. Because the accuracy of the term frequency relies on two elements, i.e. length of the document and count of the documents and terms (i.e. both within documents and across the corpus), the experiment requires that the corpus is first subjected to a general Web crawl. To improve the accuracy to calculate the length of the document, the search process ignores metadata and network identity. Stemming process are applied to handle word variations (e.g. work, works, worked, working and –work) to ensure term count accuracy. When the expression had been refined, to generate reliable results, the process was tested using a larger controlled document corpus.

5.1.5 Document Scoring and P&R Measurement (Stage E)

The retrieval function relies on the search model, which formalises the notion of relevance and provides the criterion to score and rank a document. Clearly the performance of an IR system is directly determined by the quality of the scoring function adopted. The details of different search mode scoring algorithms are demonstrated on the section 4.5. Document contents are then pattern-matched against the search term and the SLM algorithm is used to calculate page relevance rankings and subsequent search effectiveness performance comparisons.

The TREC corpus includes a set of relevance judgements for each topic, i.e. it lists a pool of relevant and non-relevant documents across the entire document collection. By referring to the judgement sets, it was possible to flag the retrieved relevant documents to calculate the P&R search effectiveness measures. Comparisons between different searching modes are outlined in the Table 4 and demonstrated using P&R graphs containing precision-recall curves for each search model executed. Precision-recall curves are determined by measuring the cumulative number of documents retrieved, and the numbers that are deemed are relevant. Graphs will be illustrated in the format shown in Fig 33, i.e. showing scales 0-100% for both precision (y-axis) and recall (x-axis). Cumulative percentage precisions are calculated at each incremental 10% interval of recall and demonstrate in the table under the P&R graph.



	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
Keyword_only	83%	67%	56%	51%	40%	35%	32%	25%	18%	14%
Mode 1	63%	59%	44%	44%	37%	39%	37%	30%	24%	19%
Mode 2	100%	91%	88%	67%	49%	50%	42%	39%	33%	27%

Fig. 33 Graph format for P&R measures

For the purely hypothetical, example P&R graph demonstrated in Fig 33, the corresponding data is shown in Table 4. It assumes that there are 50 relevant documents in the corpus, i.e. column (a). Column (b) demonstrates the search results of keyword-only, i.e. the baseline of the experiments. In column (c), the first line of Mode 2 (10% recall, 10% interval of curve in the P&R graph) has resulted in the first 5 relevant documents being retrieved in the top 5 of hit list; therefore the precision at 10% recall is 100%. However, at 20% recall point (cumulative 10 relevant documents found), the cumulative ranked document returned were 11, resulting in 91% precision at 20% recall. Ideally, the most successful outcome would be 100% precision at each recall point. However, search engines can return thousands of potential documents; therefore such an optimal state would be extremely hard to achieve in reality.

Table 4 Example of LMST P&R data.

	(a)	(b)				(c)	
%R Point	**Cumulative relevant docs. Returned**	**Cum. Docs. Returned keywords only**	**Keyword %P**	**Cum. Docs. Returned Mode1**	**Mode 1 %P**	**Cum. Docs. Returned Mode2**	**Mode 2 %P**
10%	5	6	83%	8	63%	5	100%
20%	10	15	67%	17	59%	11	91%
30%	15	27	56%	34	44%	17	88%
40%	20	39	51%	45	44%	30	67%
50%	25	63	40%	68	37%	51	49%
60%	30	85	35%	76	39%	60	50%
70%	35	110	32%	94	37%	83	42%
80%	40	157	25%	134	30%	103	39%
90%	45	250	18%	184	24%	135	33%
100%	50	370	14%	270	19%	184	27%

In general circumstances, precision is more important than recall. A typical Web user might consider the first few search pages to be relevant (high precision) and be less interested in knowing that every relevant document is retrieved. Therefore, search performance evaluation experiments will primarily consider precision outcomes in the low recall intervals, i.e. average precision value (APV) of first three recall intervals (10%, 20% and 30%). The success of the enhanced search model experiments will be determined by comparing the corresponding search mode P&R curves against the baseline P&R curve profile (Keyword-only with Dirichlet smoothing). Compared to the baseline, Fig 33 demonstrates both successful and unsuccessful search mode outcomes. In the example, the first three columns provide APV data examples.

Keyword = (83% + 67% + 56%) / 3 = 69% APV.

Mode 1 = (63% + 59% + 44%) / 3 = 55% APV.

Mode 2 = (100% + 91% + 88%) / 3 = 93% APV.

The APV results demonstrate that mode 2 has been wholly successful; the curve is consistently above other two curves and achieved best APV, i.e. 93%. Mode 1 is considered unsuccessful as APV is lower than the baseline, despite having higher precision than keyword-only beyond 60% recall.

In chapter 6, it will show that precision values can fluctuate along the recall axis. To provide a consistent approach in comparing the precision-recall curves, an average of the overall 10 query's APVs (AAPV) will be used in performance evaluations. The individual query search APV results list in Appendix C.

5.1.6 LMST Interface

The LMST interface has been implemented using the basis of George's SemSeT (Semantic Search Tool) framework (George, 2010), which was developed to exploit the vector space model and semantics-based Web search using OQE. In similar vein, it was not judged practical and feasible to use an existing commercial search engine, for three reasons.

1. A search engine page hits would depend on their own relevance algorithm;
2. Their processes do not incorporate OQE and ODC;
3. There is no capability to embed OQE and ODC in their processes.

Therefore, an independent search tool was required for the search effectiveness experiments. The interface is shown in Fig 34, and it includes three main components.

- **Search setup:** search context, keyword, search mode selection and document classification selection are in panel [i] bounded by the dashed line.
- **Process response:** this involves context choices, ontology traversal listings, OQE term results and searching process statement. These are conducted in panel [ii].
- **Search outcomes:** a ranked document search results list, output in panel [iii].

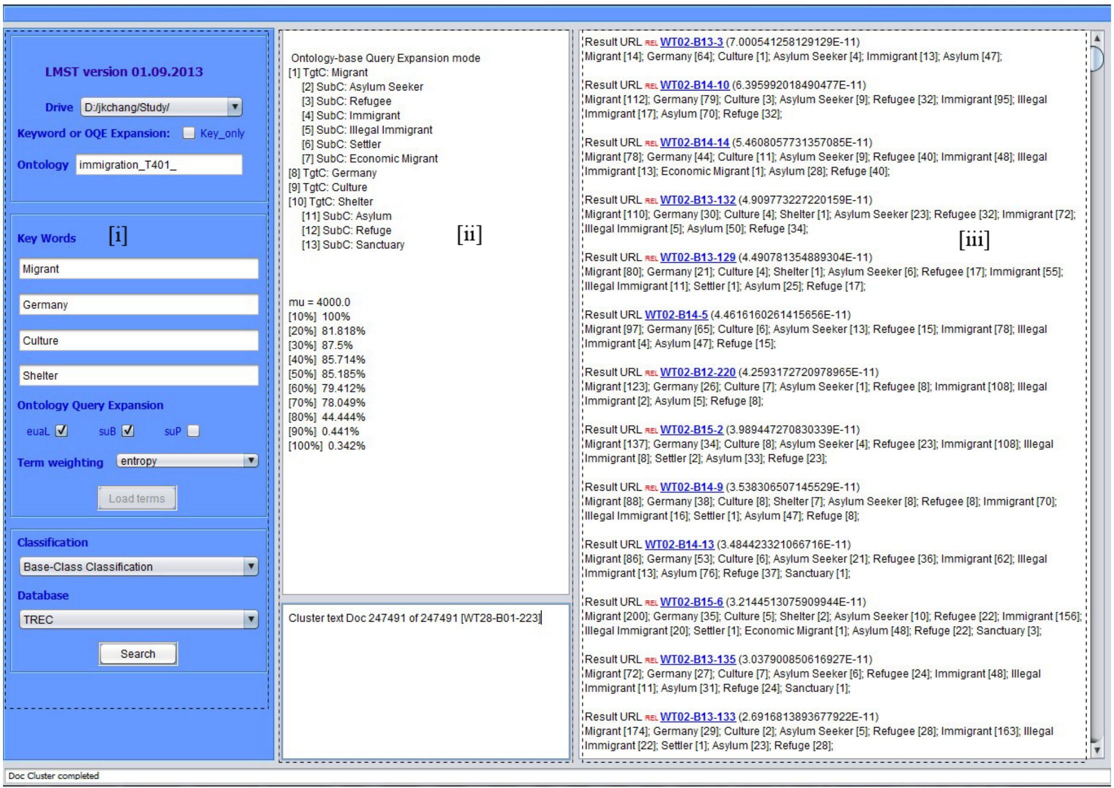


Fig. 34 The LMST interface components

5.1.7 Making a Search

This subsection represents a typical LMST search process using the example of T401, Germany immigration. First, the user can select a query term processing mode (keyword-only/OQE); if the user does not select "key_only" for keyword only search, the input box of ontology context is active and enabled for input. During Ontology input box entry, an adaptive text process will list all available Ontologies in the panel [ii].

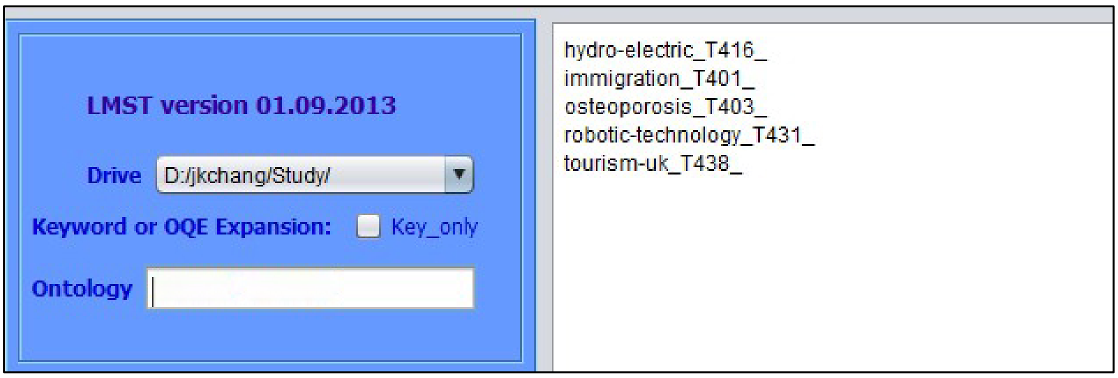


Fig. 35 Displaying all available ontologies

The ontology context choices, are accessible in the system and are determined by the Ontology input box, e.g. based on inputting "i", the immigration context has been shown, as in Fig 36.

Fig. 36 Targeting a ontology used for search

After ontology selection, LMST requires input of four search query terms. The adaptive text input process will traverse the whole ontology and list all of its concepts and individuals in panel [ii] – as shown in Fig 37.

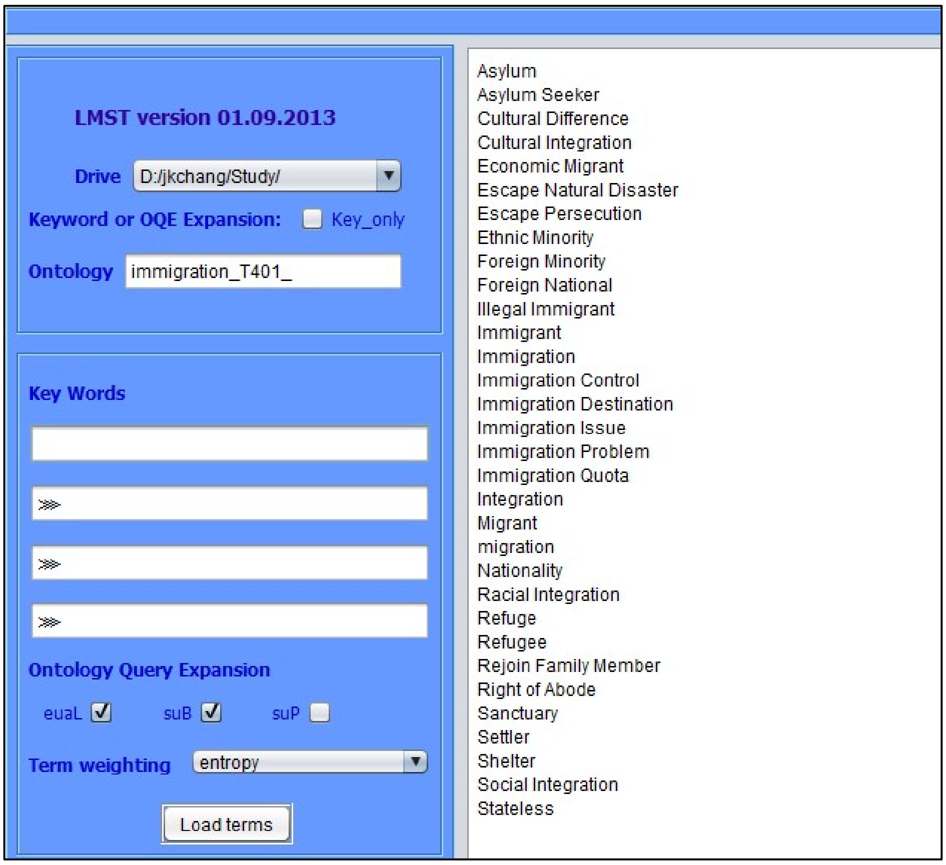


Fig. 37 Candidate query term classes for immigration context

Fig 38 shows character "m" has identified terms matching the leading character(s), i.e. concepts migrant and migration. The ontology traversal process is repeated for all 4 query terms input and if an ontology class is not matched against the text input, the user's input is accepted, as mentioned in subsection 5.1.1. Different OQE options can be chosen, e.g. subclass OQE or superclass OQE, and the term weighting approach to be used for controlling OQE process, which are discussed in subsection 4.5.1.

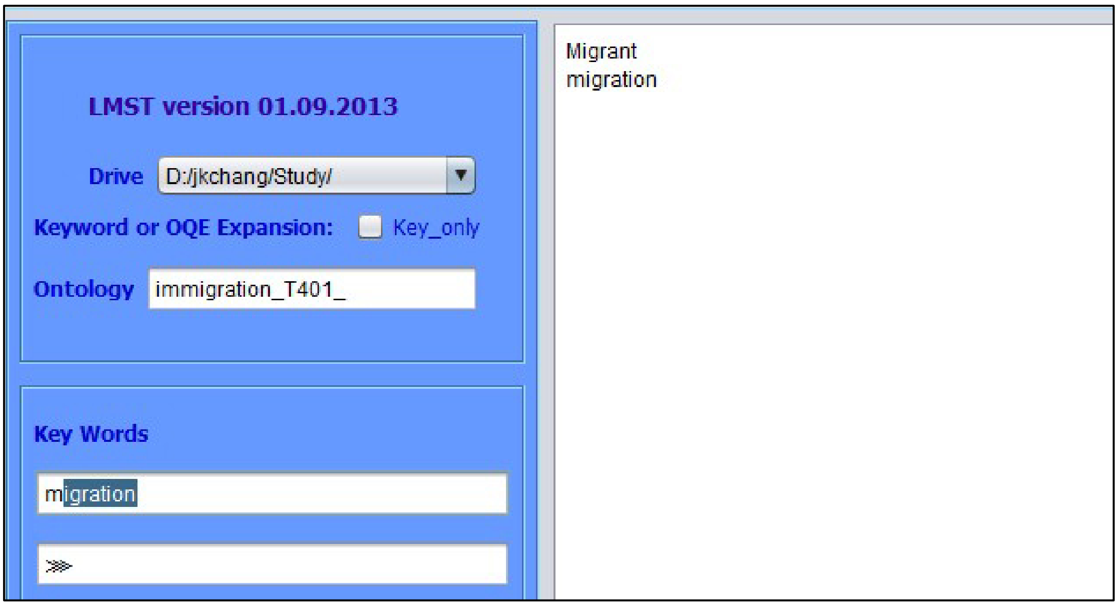


Fig. 38 Class migration selected as query term for searching

Fig. 39 shows the four query inputs for OQE process, i.e. Migrant, Germany, Culture and Shelter, in which "Germany" and "Culture" are not classes or individuals of the Immigration ontology. The input keywords are loaded by selecting "Load Terms" and the expanded, full query term set is listed in the feedback panel, i.e. each keyword is followed by its corresponding expanded terms.

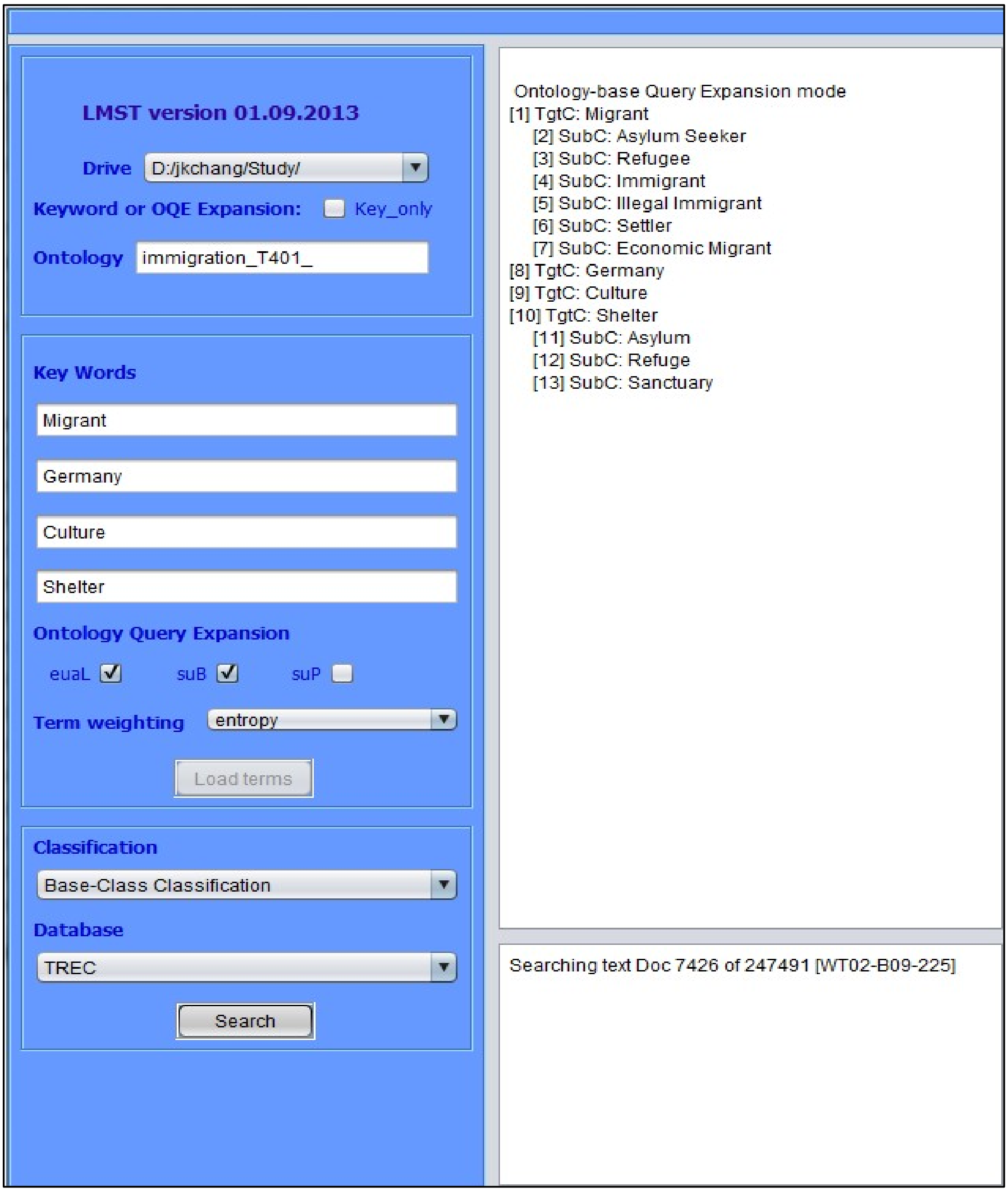


Fig. 39 Query term set generated and start for searching

After the OQE process and selection of classification approach, LMST will start searching by using the "Search" button. The search progress will show in the bottom output box of panel [ii]. – As shown in Fig 40.

Retrieved documents and scoring, together with the matched terms, are then output in ranked order in output panel [iii]. It should be pointed out that, for the purpose of the experiment, LMST shows relevance and term retrieval information; the relevant documents matched with the relevance judgement will be denoted with red flag "rel" and the tool will display the query precision result, for each 10% recall point, in the top panel [ii], as shown in Fig 40.

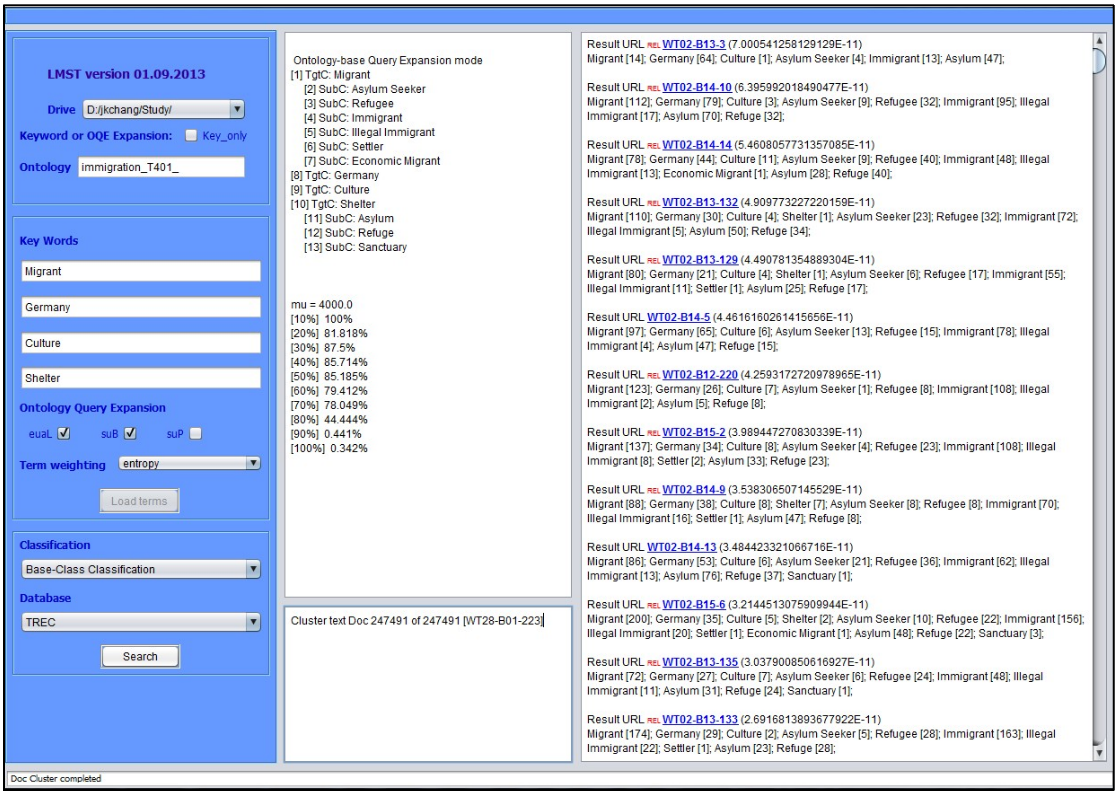


Fig. 40 LMST relevance document ranking outputs

5.2 Experiment implementation

This section will discuss the construction of each ontology and query matrix used for the search performance experiments. The TREC WT2g text corpus contains 50 topics, 5 of which were selected for the search experiments. Some topic-related modular ontologies used for the OQE and ODC processes were manually refined from existing related ontologies. Some of the concepts across the ontologies were also created according to the information retrieved from relevant Websites. A number of ontology contexts were created and the ontology metrics are shown in Table 5.

Table 5 List of created ontology

ID	Topic	Ontology	No. of classes	Maximum depth	No. of single classes
T401	Foreign Minorities, Germany	immigration	32	2	11
T403	Osteoporosis	osteoporosis	73	4	3
T416	Three Gorges Project	hydro-electric	53	2	14
T431	Robotic Technology	robotic	95	4	9
T438	Tourism, Increase	tourism	171	5	23

The query terms were formulated based on the corresponding topic statement query guidelines and the expansion-enabling ontology. As mentioned previously in subsection 5.1.1, a query contains four keywords/phrases; this query term selection approach was chosen to simulate how keywords might be applied when a Web user searches for relevant information. The query formulation also assumed that query terms would include at least one ontology class – to secure the greatest or optimal query expansion.

5.2.1 T401 "Foreign Minorities, Germany"

Fig 41 shows the T401 topic statement used for ontology and query generation. The topic description is "What language and cultural differences impede the integration of foreign minorities in Germany?" The relevant document is related to Germany and focuses on the cause of the lack of integration. Immigration difficulties and immigration problems in other countries are not relevant.

<num> Number: 401
<title> foreign minorities, Germany

<desc> Description:
What language and cultural differences impede the integration of foreign minorities in Germany?

<narr> Narrative:
A relevant document will focus on the causes of the lack of integration in a significant way; that is the mere mention of immigration difficulties is not relevant. Documents that discuss immigration problems unrelated to Germany are also not relevant.

Fig. 41 Topic statement of T401 Foreign Minorities, Germany

5.2.1.1 T401 Immigration Ontology Context

The T401 ontology is refined according to George's immigration ontology (George, 2010) which was developed for T401 and applied in query expansions to support vector space model based document search ranking. Google and immigration related Web sites (The Home Office Border Agency[2]) were also used to identify Immigration ontology context.

The *Immigration* ontology context was the smallest ontology (32 classes) of five ontologies used for the experiment, and it is the only one which is self-invented. It has a shallow hierarchy (2 layers), with 11 single classes having neither subclass nor superclass, which limits the potential for query expansion. An extract of Immigration ontology's class hierarchy is shown in Fig 42.

[2] http://www.ukba.homeoffice.gov.uk/

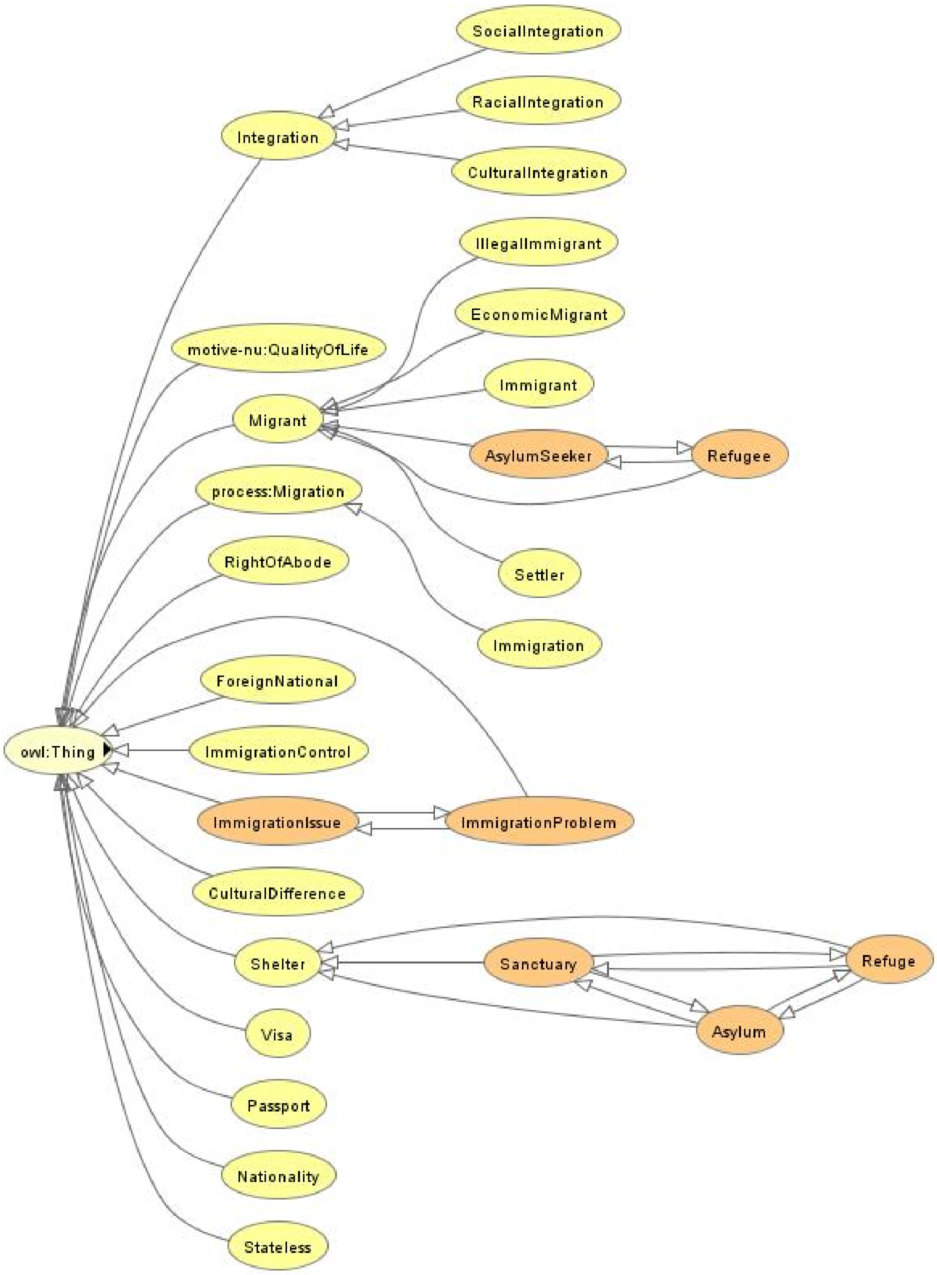


Fig. 42 Extract of the Immigration ontology context

5.2.1.2 T401 Query matrix

The query matrix was created based on the T401 topic statement query guidelines and Immigration ontology context. T401 relevant documents focus on the cause of lack of integration particular in Germany. Therefore, the "Germany" becomes a must have term in every query set. The matrix includes 19 terms across 10 query comparison sets. Table 7 illustrates the query term combinations, which were attempted to reflect the topic description.

Table 6 TREC 401 Foreign Minorities query matrix.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
asylum		●		●					●	
culture					●					●
economic migrant							●			
employment			●							
foreign national					●					
Germany	●	●	●	●	●	●	●	●	●	●
immigrant			●							
immigration				●					●	
immigration issue						●				
integration										●
language	●							●		
migrant		●			●					●
migration								●		
policy			●	●						
protection	●	●					●			
refugee						●			●	
security							●			
shelter	●							●		
visa						●				

5.2.2 T403 "Osteoporosis"

Osteoporosis is a progressive bone disease and Fig 43 shows the topic statement guidelines used to create an Osteoporosis context ontology. T403's topic description is "Find information on the effects of the dietary intakes of potassium, magnesium and fruits and vegetables as determinants of bone mineral density in elderly men and women thus preventing osteoporosis (bone decay)." Relevant documents will cover the prevention of osteoporosis, the disturbance of nutrition and mineral metabolism that results in a decrease in bone mass.

<num> Number: 403
<title> osteoporosis

<desc> Description:
Find information on the effects of the dietary intakes of potassium, magnesium and fruits and vegetables as determinants of bone mineral density in elderly men and women thus preventing osteoporosis (bone decay).

<narr> Narrative:
A relevant document may include one or more of the dietary intakes in the prevention of osteoporosis. Any discussion of the disturbance of nutrition and mineral metabolism that results in a decrease in bone mass is also relevant.

Fig. 43 Topic statement of T403 Osteoporosis

5.2.2.1 T403 Osteoporosis Ontology Context

The topic specific, modular ontology context *Osteoporosis* is a refinement of three medical ontologies, i.e. Systematized Nomenclature of Medicine Ontology[3], Bone Dysplasia Ontology[4] and Human Disease Ontology[5]. All the concepts and relationships related to "osteoporosis" (superclasses, subclasses and relationships) have been manually extracted from existing ontologies, reasoner were also applied to ensure the ontology consistency.

[3] http://bioportal.bioontology.org/ontologies/SNOMEDCT
[4] http://bioportal.bioontology.org/ontologies/BDO
[5] http://bioportal.bioontology.org/ontologies/DOID

Compared with the other 4 topics, the Osteoporosis context ontology is the most specific/specialised, i.e. most of the ontology concepts represent accepted (shared) terms for a specific medical domain. The ontology has 73 glossary concepts about bone health and deep hierarchy structure compared with the Immigration ontology, and only 3 single classes cannot be used for OQE or ODC process.

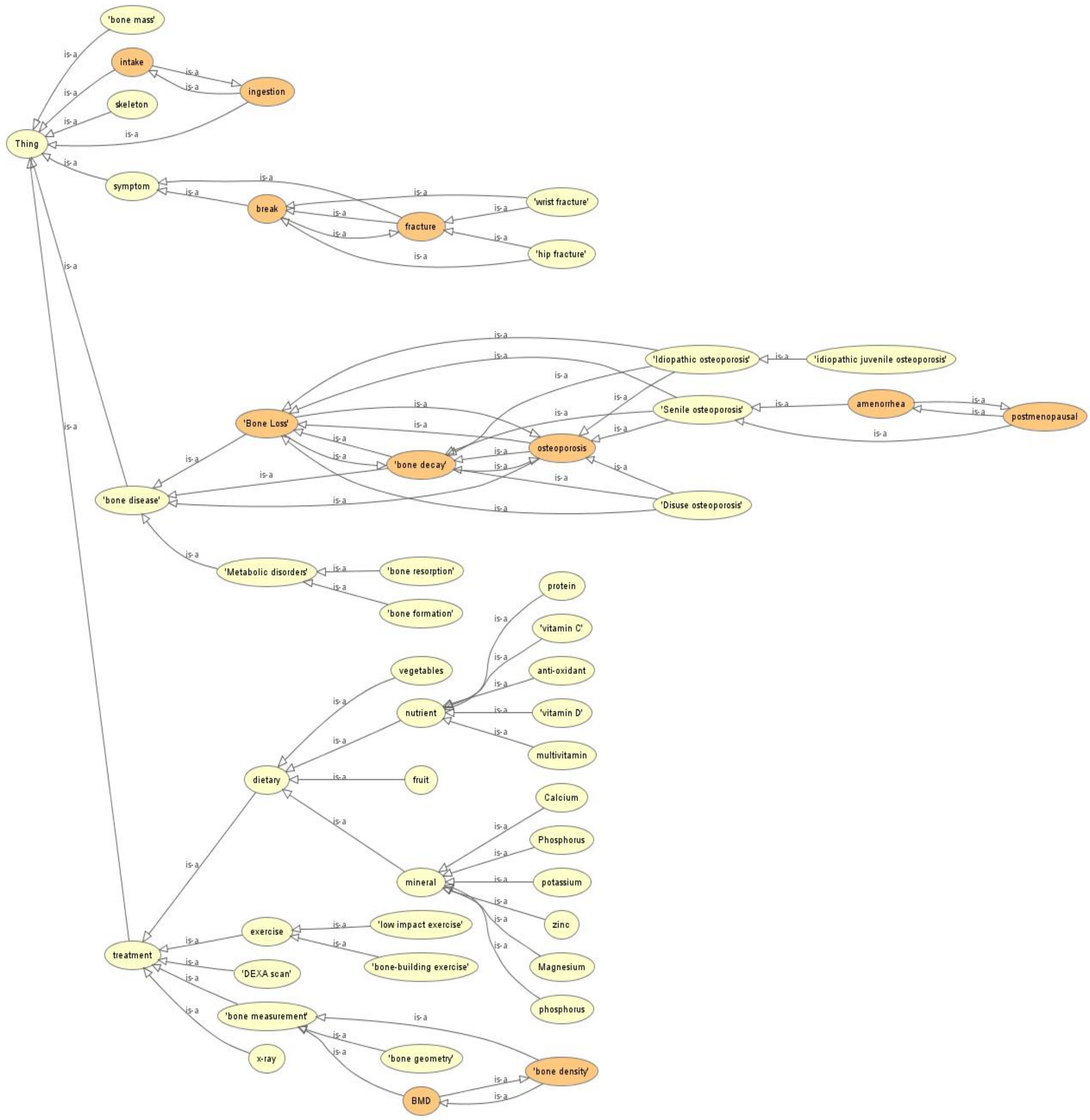


Fig. 44 Extract of the Osteoporosis ontology context

5.2.2.1 T403 Query matrix

The T403 topic statement and developed ontology context provide the basis for the query term combinations shown in table 7. The combinations are designed to reflect the query guidelines, i.e. by focusing on dietary intakes in the prevention of osteoporosis.

As the term "osteoporosis" and its synonym "bone decay" are the primary focus to determine relevant documents, they were set as "must have" terms in the 10 query sets, i.e. osteoporosis for Q1-Q5 and bone decay for Q6-Q10. The query matrix embraced 26 terms for 10 query sets.

Table 7 TREC 403 Osteoporosis query matrix

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
bone decay						●	●	●	●	●
bone disease										●
bone health						●				
bone mass										●
break							●			
density		●								
dietary	●		●						●	
elder								●		
estrogen	●									
fracture								●		
fruit									●	
hormone					●					
intake				●						
magnesium					●					
medicine	●									
metabolism				●					●	
mineral						●				
nutrient		●								
nutrition					●	●				
oestrogen			●							
osteoporosis	●	●	●	●	●					
senile			●							
treatment										●
vegetable				●				●		
vitamin		●					●			
x-ray							●			

5.2.3 T416 "Three Gorges Project"

The topic statement query guideline of T416 is shown in Fig 45. The description of this topic is "What is the status of The Three Gorges Project?" A relevant document is defined as one shows the date of project completion, estimated total cost, or the estimated electrical output. Social, political, or ecological information are not considered relevant.

<num> Number: 416 <title> Three Gorges Project <desc> Description: What is the status of The Three Gorges Project? <narr> Narrative: A relevant document will provide the projected date of completion of the project, its estimated total cost, or the estimated electrical output of the finished project. Discussion of the social, political, or ecological impact of the project are not relevant.

Fig. 45 Topic statement of T416 Three Gorges Project

5.2.3.1 T416 Hydro-electric Ontology Context

A *Hydro-electric* ontology was used for the Three Gorges Project topic's OQE and ODC enhanced semantic search process experiment. The ontology was refined from George's Hydro-electric ontology (George, 2010) and a power station ontology[6]. Two ontologies are fully merged by using Protégé, and Pellet was utilised to remove the duplicated concepts. The British Dam Society[7] and Wikipedia's Three Gorges Dam[8] Web page contents also provided a reference point to develop Hydro-electric concepts. As it has shallow class hierarchy and relatively few classes, the ontology limits query expansion capability, compared with the other 4 ontologies employed.

An extract of Hydro-electric ontology is shown in Fig 46. The complete ontology with

[6] http://ontologydesignpatterns.org/ekp/owl/PowerStation.owl
[7] http://www.britishdams.org/about_dams/hydroelectric.htm
[8] http://en.wikipedia.org/wiki/Three_Gorges_Dam

53 concepts is provided in Appendix B.

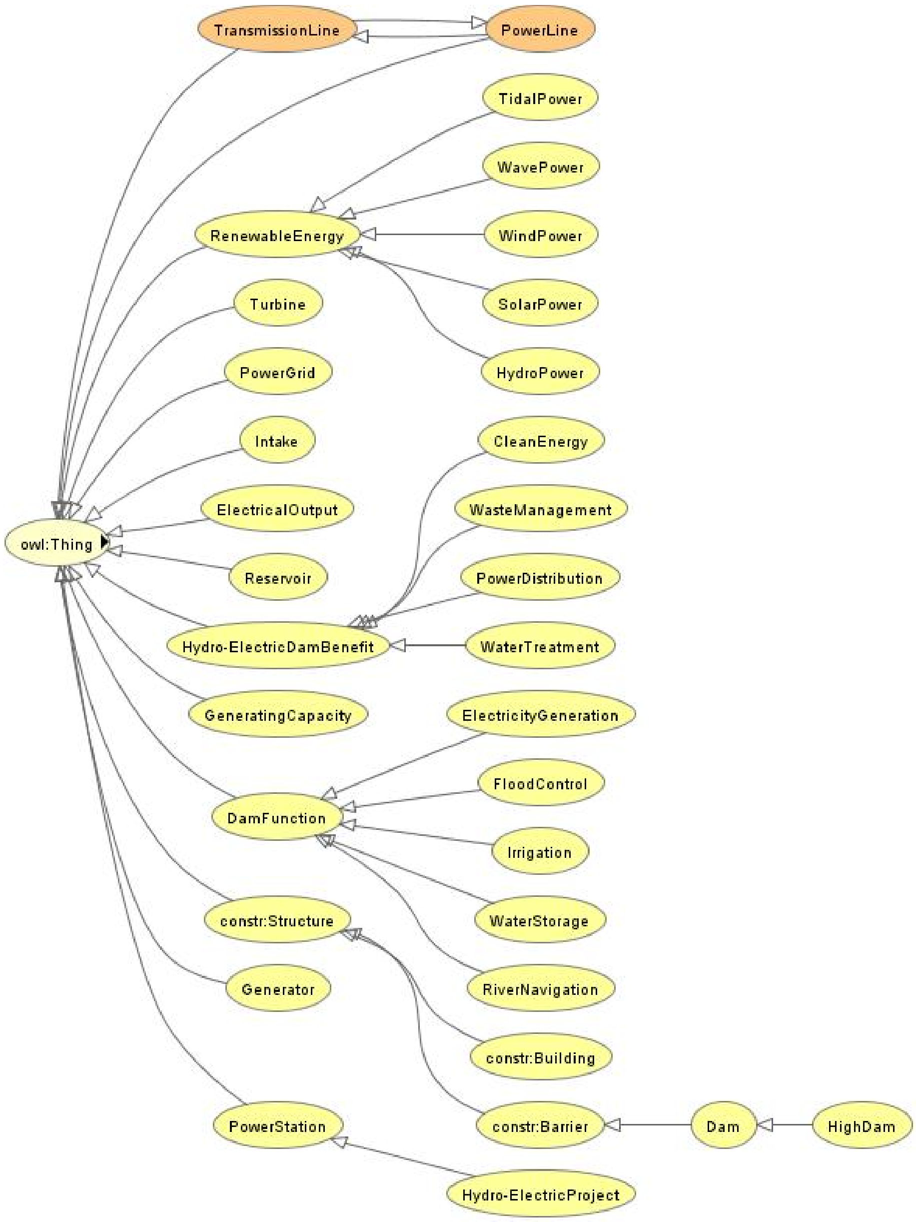


Fig. 46 Extract of the Hydro-electric ontology context

5.2.3.2 T416 Query matrix

Table 8 shows the T416 query matrix based on the TREC topic description and Hydro-electric ontology context. The matrix used 18 query terms over 10 query sets to reflect the T416 query guidelines, i.e. the focus should be on the three gorges estimated total cost, project completion date, or the estimated electrical output. Therefore, phrase "three gorges project" and its derivative "three gorges" were used for Q1-Q5 and Q6-Q10 respectively.

Table 8 TREC 416 Three Gorges Project query matrix.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
capacity	●					●				
Chang Jiang		●								
clean energy							●			
dam design								●		
dam function			●		●		●		●	
Date							●			
generator		●			●					
government								●		
hydropower	●							●		●
Intake				●						●
power grid					●					
power station	●			●						
renewable energy		●				●			●	
three gorges	●	●	●	●	●					
three gorges project						●	●	●	●	●
total cost			●						●	
Turbine				●		●				●
Yangtze			●							

5.2.4 T431 "Robotic Technology"

The topic statement of T431 is shown in Fig 47, and the topic description is "What are the latest developments in robotic technology?" The guidelines indicate that the relevant document will contain applications of robotic technology. Robotic research and simulation of robots are not considered relevant to the search.

<num> Number: 431 <title> robotic technology <desc> Description: What are the latest developments in robotic technology? <narr> Narrative: A relevant document will contain information on current applications of robotic technology. Discussions of robotics research or simulations of robots are not relevant.

Fig. 47 Topic statement of T431 Robotic Technology

5.2.4.1 Ontology Context for T431 "Robotic Technology"

It should be noted that TREC WT2g was created in 1999, therefore relevant documents should, ideally, focus on robotic applications during that period. T431 ontology was not designed specifically for T431 but used as it was associated with robotic technology.

The *Robotic Technology* ontology was manually refined according to OpenRobots Common Sense Ontology (ORO)[9], which is produced for robotic knowledge management. Meanwhile, Google was used to identify ontology sources, and a primary source was the glossary of terms[10]. An extract of the main class hierarchy of Robotic Technology concepts is shown in Fig 48.

[9] https://www.openrobots.org/wiki/oro-ontology
[10] http://en.wikipedia.org/wiki/Glossary_of_robotics

Fig. 48 Extract of the Robotic technology ontology context

It has a deep hierarchy structure and provides the potential for more extended expansion. The complete ontology totals 95 classes and is provided in Appendix B.

5.2.4.2 T431 Query matrix

The query matrix, in Table 9, was developed based on the T431 topic query description "What are the latest developments in robotic technology?" As previously mentioned the document corpus used for the experiment was generated in 1999, therefore the relevant document focus is on the robotic application during that period. The query matrix attempts to reflect the query guidelines and combines 21 query terms over the 10 query sets.

Table 9 TREC 431 Robotic Technology query matrix.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
arc welding									●	
artificial intelligence								●		●
Automata	●									
Automation				●	●					
DAPRA					●	●				
development	●								●	
Drone				●						
feedback			●							
humanoid		●		●			●			
manufacturing robot		●						●	●	
military robot					●					
mine clearance								●		
nanobot						●				
remote control										●
Robot	●	●	●	●	●					
robot technology						●	●	●	●	
robotic application	●		●			●	●			
spot welding			●							
surveillance drone		●					●			
technology										●
welding robot										●

5.2.5 T438 "Tourism, Increase"

Fig 49 shows the T438 topic statement query guidelines used for a Tourism context ontology. The T438 query description is "What countries are experiencing an increase in tourism?" This somewhat ambiguous/vague narrative indicates that relevant documents will name a country having experienced increased tourism nationwide. A document that only discusses increased projections is not relevant. Based on the given statement, it was difficult to construct precise queries for this relatively general topic.

<num> Number: 438

<title> tourism, increase

<desc> Description:
What countries are experiencing an increase in tourism?

<narr> Narrative:
A relevant document will name a country that has experienced an increase in tourism. The increase must represent the nation as a whole and tourism in general, not be restricted to only certain regions of the country or to some specific type of tourism (e.g., adventure travel). Documents discussing only projected increases are not relevant.

Fig. 49 Topic statement of T438 Tourism, Increase

5.2.5.1 Ontology Context for T438 "Tourism, Increase"

Tourism ontology was not designed specifically for T438 but is associated with tourism and is much larger than the previous four ontologies, with 171 concepts. Because the topic is not in one specific domain, a number of ontology imports were considered and four existing tourism-related ontologies were imported to construct the *Tourism* ontology, i.e. George's tourism Ontology (George, 2010), travel ontology[11], tourism ontology[12] and e-tourism ontology[13]. Three ontologies are merged using reasoner of Protégé and manually remove the inaccurate concepts. Other primary

[11] http://protege.cim3.net/file/pub/ontologies/travel/travel.owl
[12] http://goodoldai.org/ns/tgproton.owl
[13] http://sib.deri.ie/fileadmin/documents/e-tourism.owl

sources used for ontology development were: the Rough Guide [14] and the International Ecotourism Society[15].

T438's extensive and wider multi-context ontology achieved inconclusive results. Whilst the ontology has a relatively deeper class hierarchy, which provides an opportunity for query expansion terms, the less-specific nature of the ontology classes might compromise expansion quality. A small extract of the class hierarchy is shown in Fig 50.

Fig. 50 Extract of the Tourism ontology context

14 http://www.roughguides.com/destinations
15 http://www.ecotourism.org/

5.2.5.2 T438 Query matrix

The queries of T438 were difficult to be formulated, for it is a non-specific objective and wide range description. The query term combinations attempt to seek the document naming a country that has increase in tourism in general. Consequently, a wider mix of query term combinations was applied, 25 query terms crossed 10 query sets – as shown in Table 10.

Table 10 TREC 438 Tourism query matrix.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
abroad		●								
accommodation					●		●			
city				●						
cultural tourism								●		
ecotourism	●					●				●
foreign country		●			●					
holiday	●									
hotel rating						●		●		
increase			●							
museum										●
overseas tourist					●					
resort	●									
shopping							●			
sports				●						
tourism		●					●			
tourism activity	●			●						●
tourism destination					●	●				
tourism industry			●							
tourism organisation									●	
tourist							●	●	●	
town			●						●	
transportation				●						
travel								●		
travel agent			●						●	
vacation		●				●				●

6. RESULTS

This chapter details the results for the five TREC retrieval experiments. Each topic section provides comparisons between pure SLM and 53 different ontology enhanced search modes. Every search mode consists of three parts, i.e. a query process (upper case), weighting process (subscript), and adopted smoothing methods (lower case). For example, the pure SLM as the baseline is non-weighted keyword-only with Dirichlet smoothing, which can be represented as KO+entire. *KO* indicates non-weighted keyword only searching and *entire* for smoothing the document, based on the term probabilistic distribution of the entire collection.

In order to separately evaluate OQE effectiveness, term weighting and ODC to SLM, the 53 established search modes were divided into four groups, i.e. baseline versus OQE-only enhanced modes, baseline versus ODC-only enhanced modes, baseline versus weighting adjusted SLM and baseline versus the other 44 combinations of three search options. Table 11 lists all search modes in four comparison groups.

Table 11 List of approaches would be compared with the baseline

Comparison group	KO+entire		
Baseline verses OQE modes	SUB+entire	SUPER+entire	
Baseline versus ODC modes	KO+base	KO+sub	
	KO+super	KO+separate	
	KO+separate_FC		
Baseline versus weighting modes	KO_e+entire	KO_idf+entire	
Baseline versus OQE, weighting and ODC enhanced modes		SUB+base	SUB+super
	SUB_e+entire	SUPER+base	SUPER+super
	SUPER_e+entire	KO_e+base	KO_e+super
	SUB_idf+entire	SUB_e+base	SUB_e+super
	SUPER_idf+entire	SUPER_e+base	SUPER_e+super
		KO_idf+base	KO_idf+super
	SUB+separate_FC	SUB_idf+base	SUB_idf+super
	SUPER+separate_FC	SUPER_idf+base	SUPER_idf+super
	KO_e+separate_FC	SUB+sub	SUB+separate
	SUB_e+separate_FC	SUPER+sub	SUPER+separate
	SUPER_e+separate_FC	KO_e+sub	KO_e+separate
	KO_idf+separate_FC	SUB_e+sub	SUB_e+separate
	SUB_idf+separate_FC	SUPER_e+sub	SUPER_e+separate
	SUPER_idf+separate_FC	KO_idf+sub	KO_idf+separate
		SUB_idf+sub	SUB_idf+separate
		SUPER_idf+sub	SUPER_idf+separate

Given that the greatest number of search modes are in the fourth comparison group, the results of 44 search modes will not be discussed in detail and will be merged into an overall search mode discussion.

The comparisons will determine whether or not ontology enhanced mode P&R search profiles give better results. As subsection 5.1 discussed, individual query set comparison s will be based primarily on APVs at the 10-30% recall intervals - the individual query results are shown in Appendix C. In this chapter, the success of each of the TREC OQE experiments will be determined by comparing the overall average of the 10 query APV's (AAPV) outcomes against the baseline.

It should be also noted that, because of the relatively low precision levels of T403, T431 and T438, the y-axes have been reduced for better visual understanding, i.e. scales will be dependent on maximum precision values returned.

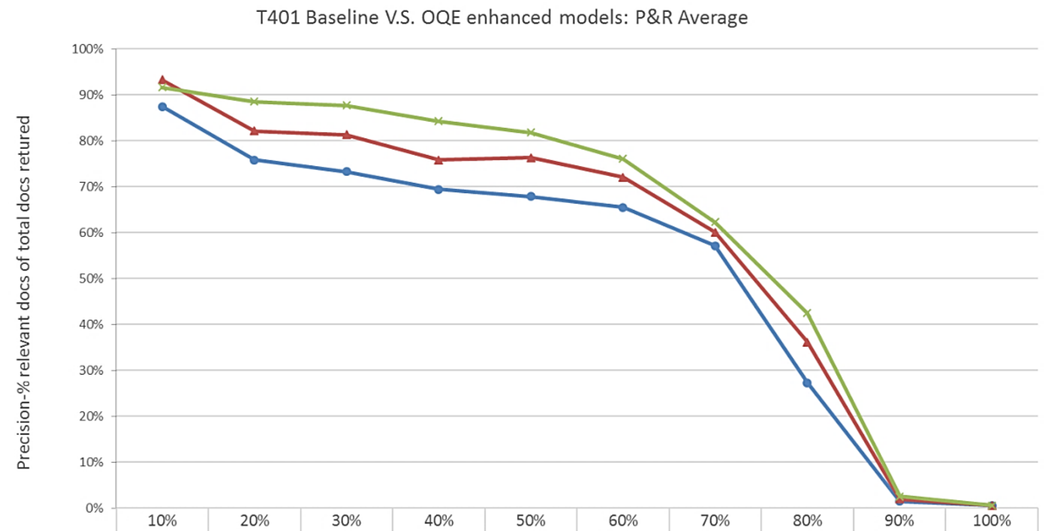
6.1 T401 "FOREIGN MINORITIES" EXPERIMENT RESULTS

This section will discuss the T401 results, with 45 relevant documents targeted across a 247,491 document total. The following graphs assess search effectiveness of the 53 established search modes.

6.1.1 Comparing between baseline with OQE query modes

This subsection compares the combined results for the baseline and two OQE search modes (SUB+entire, SUPER+entire) across 10 query sets that use the same smoothing method, i.e. Dirichlet smoothing. The P&R graph comparing the combined AAPV results for the 3 modes is shown in Fig. 51.

Based on the average number of the relevant documents returned, the primary results are that the AAPV for KO+entire was 78.89%, with 85.60% for SUB+entire and 89.32% for SUPER+entire. OQE modes provided a good AAPV improvement in search effectiveness over KO. The recall level for all three modes is 100%.

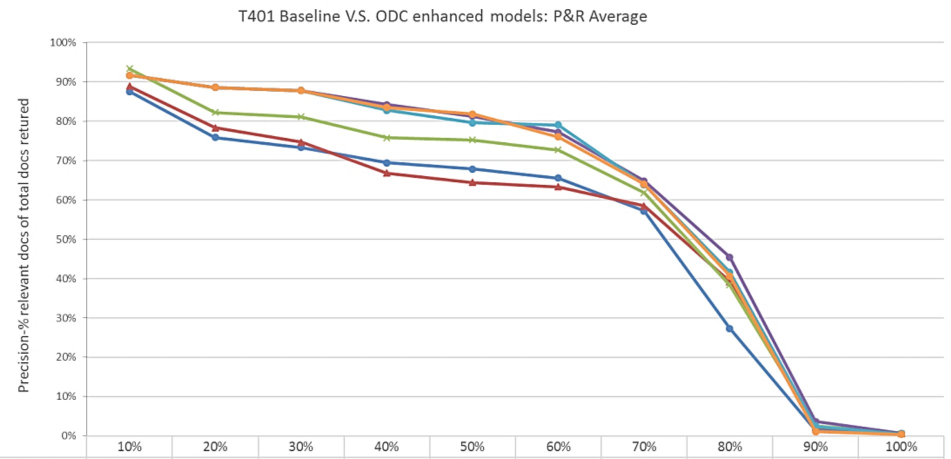


	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	87.50%	75.84%	73.33%	69.43%	67.86%	65.51%	57.19%	27.34%	1.52%	0.61%
sub+ENTIRE	93.33%	82.19%	81.29%	75.85%	76.36%	72.06%	60.10%	36.19%	2.02%	0.57%
super+ENTIRE	91.67%	88.55%	87.76%	84.26%	81.86%	76.12%	62.32%	42.51%	2.58%	0.58%

Fig. 51 T401 overall average P&R for baseline and OQE enhanced model

6.1.2 Comparing between baseline with ODC modes

To assess ODC mode effectiveness, the baseline (KO+entire) was compared with five ODC modes, i.e. KO+base, KO+sub, KO+super, KO+separate and KO+separate_FC. The following graphs show overall P&R comparisons for all 10 query sets.



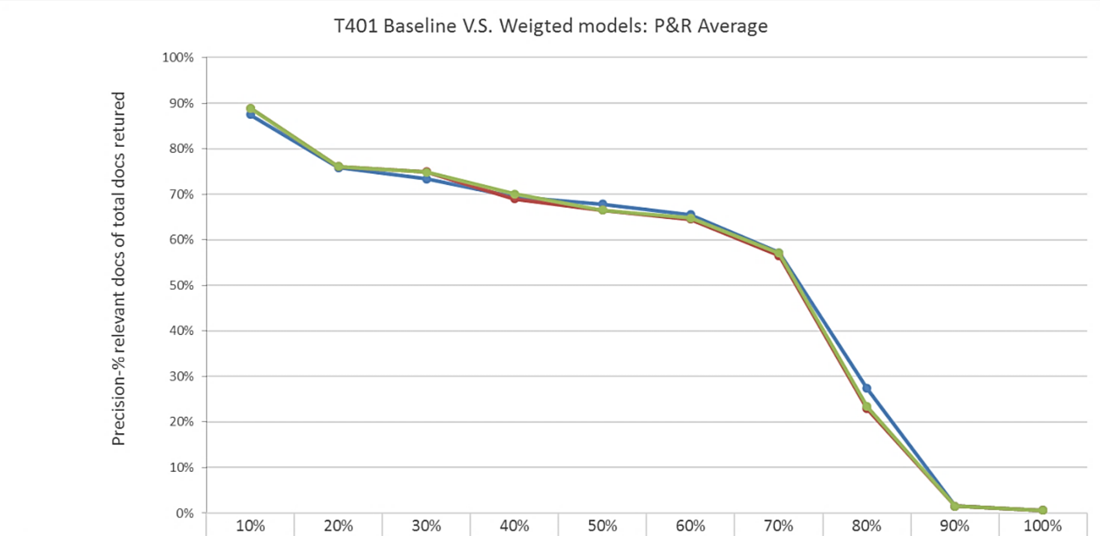
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	87.50%	75.84%	73.33%	69.43%	67.86%	65.51%	57.19%	27.34%	1.52%	0.61%
ko+BASE	88.85%	78.34%	74.65%	66.82%	64.44%	63.29%	58.46%	39.28%	2.28%	0.62%
ko+SUB	93.33%	82.19%	81.14%	75.74%	75.18%	72.70%	61.80%	38.36%	2.64%	0.64%
ko+SUPER	91.67%	88.55%	87.76%	84.23%	81.29%	77.21%	64.80%	45.42%	3.70%	0.66%
ko+SEPARATE	91.67%	88.55%	87.76%	82.80%	79.56%	79.02%	63.86%	41.66%	2.44%	0.58%
ko+SEPARATE_FC	91.67%	88.55%	87.76%	83.48%	81.84%	76.05%	64.02%	40.68%	1.11%	0.43%

Fig. 52 T401 overall average P&R for baseline and ODC enhanced model

Fig. 52 illustrates the search effectiveness results of the combined 10 query sets for baseline and ODC modes. It shows that applying ODC to SLM achieved an improvement, with 78.89% for KO+entire, 80.61% for KO+base, 85.55% for KO+sub, 89.32% for KO+super and 89.32% for both KO+separate and KO+separate_FC. Despite there being no difference between last two ODC enhanced modes, search effectiveness has been improved.

6.1.3 Comparing between non-weighted with weighted modes

This group evaluates the effect, of introducing term weighting into the pure language model, on search effectiveness. Two approaches (entropy and tf-idf) have been adopted, to adjust query weighting. The graph in Fig 53 illustrates overall P&R comparison for all 10 query sets.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	87.50%	75.84%	73.33%	69.43%	67.86%	65.51%	57.19%	27.34%	1.52%	0.61%
ko_entropy+ENTIRE	88.85%	76.11%	74.93%	68.98%	66.49%	64.55%	56.55%	22.92%	1.51%	0.65%
ko_idf+ENTIRE	88.85%	76.11%	74.84%	70.03%	66.52%	64.79%	57.07%	23.44%	1.54%	0.65%

Fig. 53 T401 overall average P&R for baseline and weighting adjusted model

The results show there is negligible difference between the non-weighted baseline and weighted SLM, i.e. 78.89% for baseline, 79.96% for entropy weighted mode and 79.93% for tf-idf adjusted SLM searches. Despite the small difference between the enhanced mode and baseline mode, weighted SLM produced a better result.

6.1.4 Comparison of precision results across all search modes

Tables 12 - 14 show the average search effectiveness comparisons, for all 54 search modes, at every recall interval. The last two rows show, firstly the average precision (AAPV) over the first 30% of recall points, then secondly, the whole range of recall. Table 15 illustrates average AAPV of 54 search modes. The gradient colour, from green to red, shows the AAPV value from small to large, i.e. darker green denotes smallest value and darker red denotes largest value.

The merged experiment results indicate that ontology enhanced search models achieved varying degrees of search improvement, against baseline, demonstrating that hierarchical OQE and/or term weighting and/or ontology based document classification can have a positive impact on pure SLM searching.

The best performing search mode in T401 was SUPER_e+separate, which achieved approximately 11% (89.98%-78.89%) precision improvement over the baseline. A further outcome was that all search modes retrieved the 45 relevant documents, resulting in 100% recall.

Table 16 demonstrates the relative improvement for each enhanced search mode respectively, i.e. Improvement = (AAPV of search mode-AAPV of baseline)/ AAPV of baseline. The table indicates that established search modes provide varying degrees of improved search effectiveness. SUPER_e+separate provide the highest improvement against pure language model, i.e. 14.06%.

Table 12 T401 Summary of average precision across 10 query sets of Non-weighted modes

T401	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	87.50%	88.85%	93.33%	91.67%	91.67%	91.67%	93.33%	93.33%	93.33%	91.67%	91.67%	91.67%	91.67%	91.67%	91.67%	91.67%	91.67%	91.67%
20%	75.84%	78.34%	82.19%	88.55%	88.55%	88.55%	82.19%	82.19%	82.19%	88.55%	88.55%	88.55%	88.55%	88.55%	88.55%	88.55%	88.55%	88.55%
30%	73.33%	74.65%	81.14%	87.76%	87.76%	87.76%	81.29%	81.14%	81.14%	87.76%	87.76%	87.76%	87.76%	87.76%	87.76%	87.76%	87.76%	87.76%
40%	69.43%	66.82%	75.74%	84.23%	82.80%	83.48%	75.85%	75.74%	75.74%	84.23%	82.80%	83.48%	84.26%	82.09%	82.09%	84.23%	82.80%	83.48%
50%	67.86%	64.44%	75.18%	81.29%	79.56%	81.84%	76.36%	75.18%	75.18%	81.29%	79.56%	81.84%	81.86%	80.95%	80.95%	81.29%	79.56%	81.84%
60%	65.51%	63.29%	72.70%	77.21%	79.02%	76.05%	72.06%	72.70%	72.70%	77.21%	79.02%	76.05%	76.12%	77.21%	77.21%	77.21%	79.02%	76.05%
70%	57.19%	58.46%	61.80%	64.80%	63.86%	64.02%	60.10%	61.84%	61.80%	64.80%	63.86%	64.02%	62.32%	64.80%	64.76%	64.80%	63.86%	64.02%
80%	27.34%	39.28%	38.36%	45.42%	41.66%	40.68%	36.19%	38.40%	38.36%	45.15%	41.63%	40.68%	42.51%	45.76%	45.71%	45.15%	41.63%	40.68%
90%	1.52%	2.28%	2.64%	3.70%	2.44%	1.11%	2.02%	2.57%	2.59%	3.65%	2.42%	1.09%	2.58%	3.43%	3.45%	3.65%	2.42%	1.07%
100%	0.61%	0.62%	0.64%	0.66%	0.58%	0.43%	0.57%	0.62%	0.62%	0.64%	0.57%	0.41%	0.58%	0.63%	0.62%	0.62%	0.55%	0.40%
Avg30	78.89%	80.61%	85.55%	89.32%	89.32%	89.32%	85.60%	85.55%	85.55%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%
AVG100	52.63%	53.70%	58.37%	62.53%	61.79%	61.56%	58.00%	58.37%	58.37%	62.49%	61.78%	61.56%	61.82%	62.28%	62.28%	62.49%	61.78%	61.55%

Table 13 T401 Summary of average precision across 10 query sets of Entropy-weighted modes

T401	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	88.85%	91.11%	95.00%	95.00%	93.33%	91.67%	95.00%	95.00%	95.00%	95.00%	93.33%	91.67%	95.00%	95.00%	95.00%	95.00%	93.33%	91.67%
20%	76.11%	77.92%	81.82%	86.91%	87.86%	87.55%	80.51%	81.82%	81.82%	86.91%	87.05%	86.73%	85.91%	87.73%	87.73%	87.73%	89.36%	86.73%
30%	74.93%	76.07%	80.04%	86.84%	87.25%	89.25%	80.82%	80.04%	80.04%	86.84%	87.25%	89.25%	88.22%	86.84%	86.84%	86.84%	87.25%	89.25%
40%	68.98%	67.88%	76.49%	82.75%	80.46%	85.36%	76.21%	76.49%	76.49%	82.75%	80.46%	85.36%	85.36%	82.75%	82.75%	82.75%	80.46%	85.36%
50%	66.49%	64.74%	74.12%	81.25%	78.94%	81.37%	75.11%	74.68%	74.12%	81.25%	78.94%	81.37%	80.82%	81.81%	81.25%	81.25%	78.94%	81.37%
60%	64.55%	62.24%	72.69%	78.92%	78.06%	80.93%	73.92%	72.89%	72.69%	78.92%	78.06%	80.93%	79.75%	79.69%	79.49%	78.92%	78.06%	80.93%
70%	56.55%	57.70%	63.72%	69.94%	67.83%	70.70%	63.72%	64.25%	63.72%	69.94%	67.83%	70.70%	68.32%	70.77%	70.23%	69.94%	67.83%	70.70%
80%	22.92%	33.78%	34.28%	41.52%	48.77%	47.50%	33.76%	34.36%	34.09%	41.32%	48.56%	47.50%	41.45%	41.48%	41.21%	41.32%	48.56%	47.50%
90%	1.51%	2.44%	2.61%	3.10%	2.71%	1.05%	2.04%	2.51%	2.53%	3.00%	2.69%	1.01%	2.60%	2.90%	2.91%	2.96%	2.68%	0.99%
100%	0.65%	0.68%	0.68%	0.66%	0.62%	0.42%	0.58%	0.64%	0.63%	0.63%	0.60%	0.39%	0.57%	0.63%	0.62%	0.61%	0.58%	0.37%
Avg30	79.96%	81.70%	85.62%	89.58%	89.48%	89.49%	85.44%	85.62%	85.62%	89.58%	89.21%	89.21%	89.71%	89.86%	89.86%	89.86%	89.98%	89.21%
AVG100	52.14%	53.46%	58.15%	62.69%	62.58%	63.58%	58.17%	58.27%	58.11%	62.66%	62.48%	63.49%	62.80%	62.96%	62.80%	62.73%	62.71%	63.49%

Table 14 T401 Summary of average precision across 10 query sets of tfidf-weighted modes

T401	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	88.85%	89.44%	95.00%	93.33%	93.33%	91.67%	93.33%	95.00%	95.00%	93.33%	93.33%	91.67%	91.67%	93.33%	93.33%	93.33%	93.33%	91.67%
20%	76.11%	77.76%	81.66%	86.91%	88.73%	86.73%	81.33%	81.66%	81.66%	86.91%	88.73%	86.73%	87.73%	86.91%	86.91%	86.91%	88.73%	86.73%
30%	74.84%	76.26%	80.02%	88.22%	86.86%	89.25%	81.76%	80.02%	80.02%	88.22%	86.86%	89.25%	87.76%	88.22%	88.22%	88.22%	86.86%	89.25%
40%	70.03%	69.03%	76.66%	83.29%	80.49%	84.93%	77.24%	75.92%	76.66%	83.29%	80.49%	84.93%	84.62%	83.29%	83.29%	83.29%	80.49%	84.93%
50%	66.52%	65.45%	75.28%	83.44%	79.67%	81.45%	74.95%	75.58%	75.28%	83.44%	79.67%	81.45%	80.65%	83.75%	83.44%	83.44%	79.67%	81.45%
60%	64.79%	62.96%	72.78%	80.32%	77.63%	80.59%	75.24%	72.78%	72.78%	80.32%	77.63%	80.59%	79.96%	80.88%	80.88%	80.32%	77.63%	80.59%
70%	57.07%	57.95%	63.99%	70.54%	67.22%	70.03%	65.09%	64.49%	63.99%	70.54%	67.22%	70.03%	68.48%	71.22%	70.73%	70.54%	67.22%	70.03%
80%	23.44%	32.85%	34.44%	42.71%	48.34%	47.36%	38.59%	34.41%	34.24%	42.50%	48.14%	47.36%	45.59%	42.74%	42.57%	42.50%	48.14%	47.36%
90%	1.54%	2.50%	2.50%	3.17%	2.73%	1.06%	2.15%	2.42%	2.43%	3.07%	2.71%	1.01%	2.75%	2.96%	2.97%	3.04%	2.70%	1.00%
100%	0.65%	0.67%	0.67%	0.66%	0.62%	0.42%	0.59%	0.64%	0.63%	0.63%	0.59%	0.39%	0.57%	0.63%	0.62%	0.61%	0.58%	0.37%
Avg30	79.93%	81.15%	85.56%	89.49%	89.64%	89.21%	85.47%	85.56%	85.56%	89.49%	89.64%	89.21%	89.05%	89.49%	89.49%	89.49%	89.64%	89.21%
AVG100	52.37%	53.49%	58.30%	63.26%	62.56%	63.35%	59.03%	58.29%	58.27%	63.23%	62.54%	63.34%	62.98%	63.39%	63.30%	63.22%	62.53%	63.34%

Table 15 T401 Average 30% APV of 54 search modes

	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Baseline	78.89%	80.61%	85.55%	89.32%	89.32%	89.32%	85.60%	85.55%	85.55%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%	89.32%
Entropy	79.96%	81.70%	85.62%	89.58%	89.48%	89.49%	85.44%	85.62%	85.62%	89.58%	89.21%	89.21%	89.71%	89.86%	89.86%	89.86%	89.98%	89.21%
Tf-Idf	79.93%	81.15%	85.56%	89.49%	89.64%	89.21%	85.47%	85.56%	85.56%	89.49%	89.64%	89.21%	89.05%	89.49%	89.49%	89.49%	89.64%	89.21%

Table 16 T401 improvements

T401	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	0.00%	2.18%	8.44%	13.23%	13.23%	13.23%	8.51%	8.44%	8.44%	13.23%	13.23%	13.23%	13.23%	13.23%	13.23%	13.23%	13.23%	13.23%
entropy	1.36%	3.56%	8.53%	13.55%	13.42%	13.43%	8.31%	8.53%	8.53%	13.55%	13.08%	13.09%	13.71%	13.90%	13.90%	13.90%	14.06%	13.09%
Tf-Idf	1.32%	2.87%	8.45%	13.43%	13.63%	13.09%	8.34%	8.45%	8.45%	13.43%	13.63%	13.09%	12.88%	13.43%	13.43%	13.43%	13.63%	13.09%
AVG	0.89%	2.87%	8.48%	13.40%	13.43%	13.25%	8.39%	8.48%	8.48%	13.40%	13.31%	13.13%	13.27%	13.52%	13.52%	13.52%	13.64%	13.13%

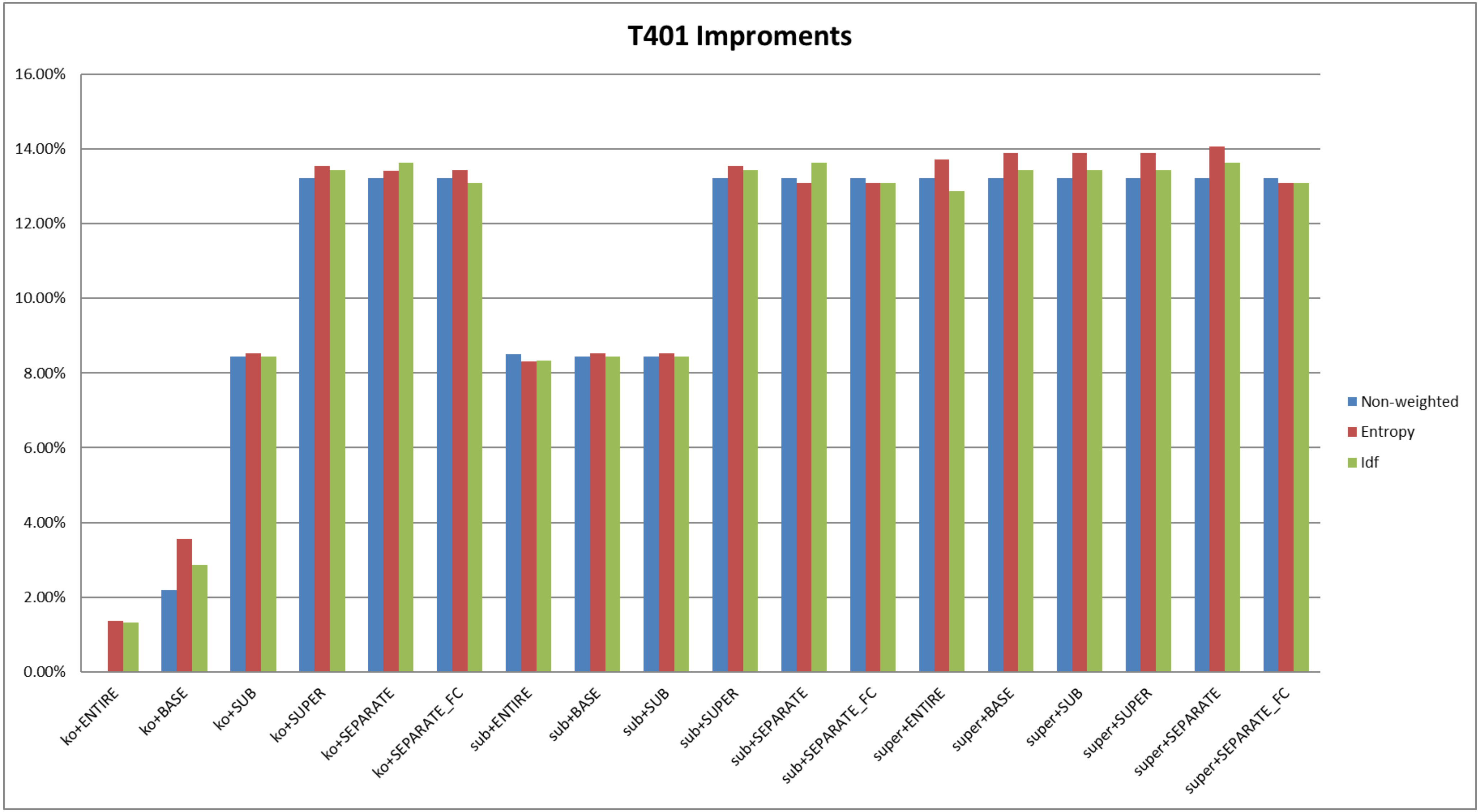


Fig. 54 T401 search mode improvements

6.1.5 Reflection on Hypotheses

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

The Immigration ontology was developed specifically for T401. The experiment results illustrated in 6.1.1 show that OQE-only enhanced search model achieved better results compared with keyword only search. The average precision for SUPER+entire was approximately 11% higher than keyword only modes (89% vs. 78%). Recall improvement was inconclusive, with each search mode achieved 100% recall.

The experiment has provided a positive indication of the search effectiveness benefits of applying OQE strategies to the pure SLM. The AAPV outcomes support the hypothesis.

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

Subsection 6.1.2 provides comparisons between the non-weighted pure language model and SLM assigned entropy or tf-idf weighting. The experiments showed positive results against the baseline, i.e. term weighting enhanced search modes made slight improvements over baseline.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or classification based local smoothing strategies are more effective than the global strategies. "*

The ontology based document classification provides the corpus structure to the language model smoothing and the results show the benefits of document classification. Superclass ODC (KO+super) provides best smoothing results, i.e. 11% precision improvement achieved against the baseline. Further, OQE modes with ODC achieved even higher precision than pure OQE modes.

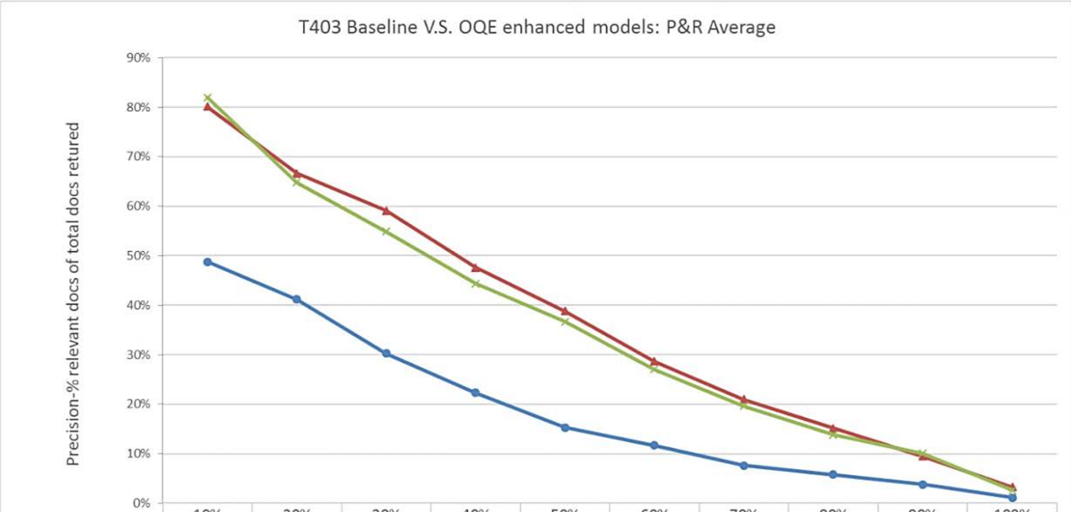
The above outcomes demonstrate that the hypotheses have been supported.

6.2 T403 "OSTEOPOROSIS" EXPERIMENT RESULTS

The T403 experiments attempted to retrieve 91 relevant documents across 247,491 documents in the corpus. Search effectiveness was measured using 10 query set combinations in 3 query processes (KO, SUB and SUPER), 3 weighting processes (non-weighting, entropy and tf-idf) and 6 smoothing modes (entire, base, sub, super, separate and separate_FC), i.e. 54 search modes in total. There were 4 search mode groups, i.e. baseline vs. OQE modes, baseline vs. ODC modes, baseline vs. weighting adjusted modes and an overall analysis. The 10 query sets results are based on overall average P&R (AAPV). P&R results for each individual query are illustrated in Appendix C.

6.2.1 Comparing between baseline with OQE query modes

Fig 55 shows merged AAPV using combined P&R results, for comparison between baseline and two OQE modes across 10 query sets, i.e. SUB+entire and SUPER+entire. The adopted Dirichlet smoothing estimates the probability distribution of unseen words in the entire collection.



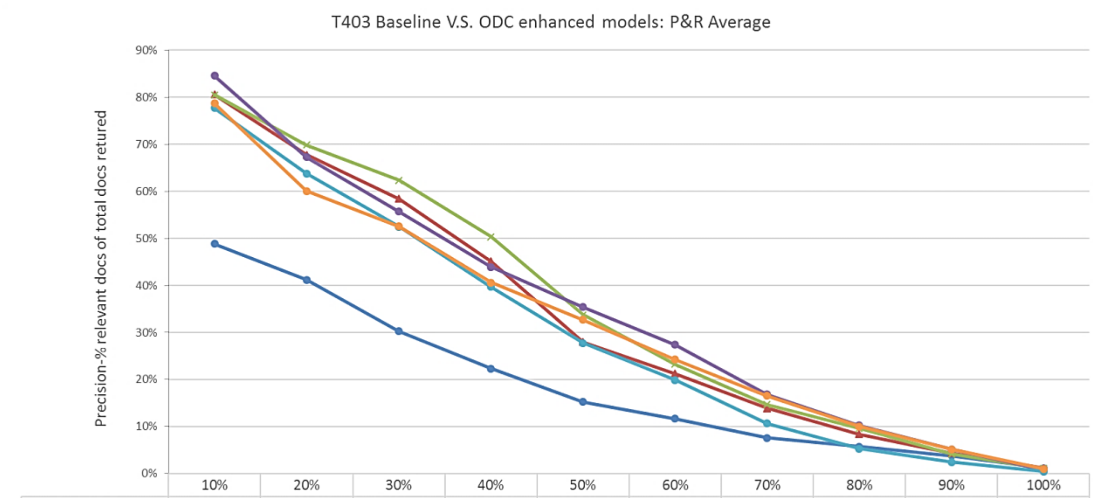
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	48.80%	41.17%	30.25%	22.32%	15.27%	11.66%	7.59%	5.76%	3.77%	1.15%
sub+ENTIRE	80.13%	66.64%	59.08%	47.58%	38.78%	28.67%	20.96%	15.11%	9.48%	3.22%
super+ENTIRE	81.97%	64.75%	54.89%	44.38%	36.70%	26.98%	19.63%	13.84%	10.03%	2.59%

Fig. 55 T403 overall average P&R for baseline and OQE enhanced model

In Fig. 55, the primary observation is that the AAPV for KO+entire was 40.07%, SUB+entire was 68.62% and SUPER+entire was 67.21%. The curves of three OQE modes show better precision results against KO at all recall points. This demonstrates the benefit of OQE; in particular, using the SUB+entire mode.

6.2.2 Comparing between baseline with ODC modes

The following P&R graphs show combined comparisons ODC mode search effectiveness. Baseline KO+entire is compared with five ODC modes, i.e. KO+base, KO+sub, KO+super, KO+separate and KO+separate_FC. The AAPV of all 10 query sets are shown in Fig. 56.



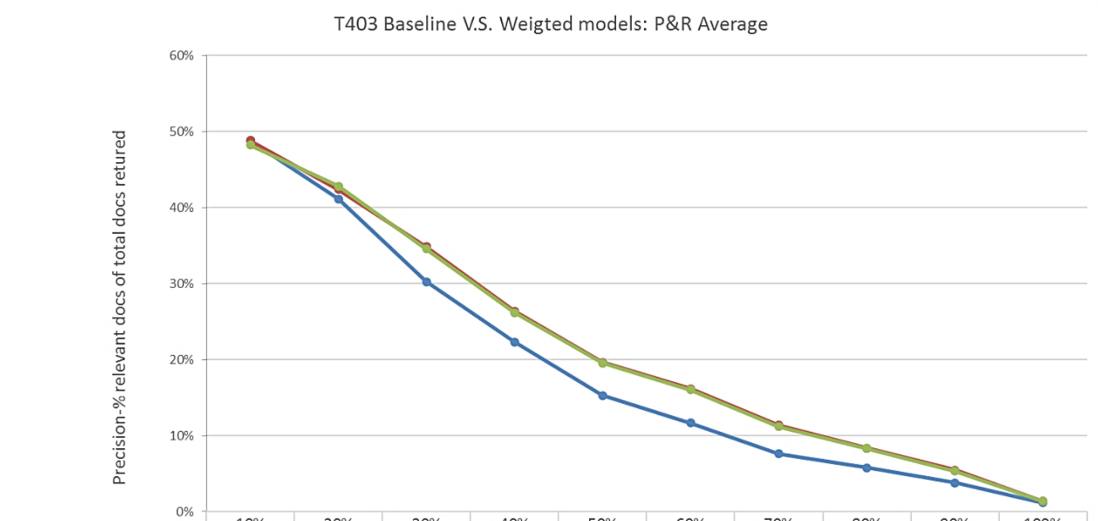
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	48.80%	41.17%	30.25%	22.32%	15.27%	11.66%	7.59%	5.76%	3.77%	1.15%
ko+BASE	80.58%	67.79%	58.38%	45.10%	27.96%	21.24%	13.89%	8.40%	4.40%	1.13%
ko+SUB	80.47%	69.84%	62.32%	50.30%	33.74%	23.28%	14.64%	9.62%	4.16%	1.06%
ko+SUPER	84.55%	67.28%	55.71%	43.93%	35.36%	27.38%	16.80%	10.20%	5.17%	0.69%
ko+SEPARATE	77.73%	63.77%	52.51%	39.70%	27.76%	19.88%	10.70%	5.33%	2.46%	0.47%
ko+SEPARATE_FC	78.72%	60.03%	52.53%	40.61%	32.67%	24.26%	16.51%	9.99%	5.14%	1.01%

Fig. 56 T403 overall average P&R for baseline and ODC enhanced model

The primary results show ODC modes produced remarkably better results over baseline, with ODC enhanced mode curves all above the baseline. The AAPV for KO+entire is 40.07%, 68.92% for KO+base, 70.88% for KO+sub, 69.18 for KO+super, 64.67% for KO+separate and 63.76% KO+separate_FC.

6.2.3 Comparing between non-weighted with weighted modes

The P&R graph (Fig. 57) compares the pure language model without term weighting and two weight adjusted modes, i.e. KO_e+entire and KO_idf+entire. The AAPV of all 10 query sets are shown, with individual query results shown in Appendix C.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	48.80%	41.17%	30.25%	22.32%	15.27%	11.66%	7.59%	5.76%	3.77%	1.15%
ko_entropy+ENTIRE	48.80%	42.36%	34.85%	26.40%	19.62%	16.16%	11.36%	8.36%	5.48%	1.35%
ko_idf+ENTIRE	48.22%	42.81%	34.54%	26.19%	19.55%	16.02%	11.17%	8.28%	5.35%	1.34%

Fig. 57 T403 overall average P&R for baseline and weighting adjusted model

The primary results show that weight-assigned mode outcomes were only marginally better than baseline, i.e. the AAPV for KO+entire was 40.07%, 42.00% for entropy weighting adjusted mode, and 41.86% for tf-idf mode. Secondary outcomes were that assigning a weighting did not always provide positive results and that tf-idf weighting produced marginally worse results at 10% recall.

6.2.4 Comparison of precision results across all search modes

This subsection details T403's merged AAPV search effectiveness results for 54 search modes. Tables 17-19 show average precision at every recall interval for all 10 query sets. The last two rows show average precisions, first at 30% recall and second, over the full recall range. As with T403, table 20 highlights value order using gradient colour, i.e.

green to red indicate values from small to large, respectively.

The primary outcomes were that extending OQE and ODC to SLM produced significant search effectiveness improvements, whereas, term weighting produced small improvements against the baseline.

Merged AAPV values of all search modes are shown in Table 20 and indicate that all search modes produced better search effectiveness compared with baseline. It again highlights the benefit of applying ontology to improve SLM search effectiveness. The best search mode in T403 was SUPER_e+separate (77.67% AAPV) which achieved approximately a 37% precision improvement over baseline (40.04% AAPV). The results also demonstrated that OQE can produce better recall.

Fig. 58 visually compares the relative search effectiveness of each enhanced search model, compared with baseline, and shows that SUPER_e+separate provided significant search improvement.

Table 17 T403 Summary of average precision across 10 query sets of Non-weighted modes

T403	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	48.80%	80.58%	80.47%	84.55%	77.73%	78.72%	80.13%	80.86%	80.97%	84.55%	77.73%	78.72%	81.97%	82.65%	82.47%	84.55%	75.97%	78.72%
20%	41.17%	67.79%	69.84%	67.28%	63.77%	60.03%	66.64%	74.21%	70.62%	68.12%	64.79%	61.48%	64.75%	72.68%	69.89%	69.50%	66.31%	61.48%
30%	30.25%	58.38%	62.32%	55.71%	52.51%	52.53%	59.08%	65.40%	63.83%	56.73%	51.99%	52.88%	54.89%	61.26%	58.64%	57.38%	51.16%	52.56%
40%	22.32%	45.10%	50.30%	43.93%	39.70%	40.61%	47.58%	56.23%	52.87%	45.45%	40.03%	42.67%	44.38%	49.63%	46.71%	45.85%	40.07%	42.81%
50%	15.27%	27.96%	33.74%	35.36%	27.76%	32.67%	38.78%	42.46%	36.70%	38.45%	28.63%	34.34%	36.70%	40.16%	38.14%	38.79%	28.86%	34.15%
60%	11.66%	21.24%	23.28%	27.38%	19.88%	24.26%	28.67%	30.20%	27.75%	29.46%	21.17%	25.92%	26.98%	28.53%	27.74%	29.55%	20.98%	25.97%
70%	7.59%	13.89%	14.64%	16.80%	10.70%	16.51%	20.96%	22.01%	21.05%	21.05%	13.97%	17.75%	19.63%	20.66%	20.29%	20.99%	13.83%	17.84%
80%	5.76%	8.40%	9.62%	10.20%	5.33%	9.99%	15.11%	15.24%	15.62%	14.68%	8.46%	13.59%	13.84%	15.09%	14.10%	14.57%	8.36%	13.44%
90%	3.77%	4.40%	4.16%	5.17%	2.46%	5.14%	9.48%	9.07%	9.77%	9.93%	5.17%	9.60%	10.03%	10.39%	9.52%	9.60%	5.26%	9.49%
100%	1.15%	1.13%	1.06%	0.69%	0.47%	1.01%	3.22%	1.82%	2.69%	2.66%	1.66%	2.34%	2.59%	1.41%	2.91%	2.56%	1.65%	2.35%
Avg30	40.07%	68.92%	70.88%	69.18%	64.67%	63.76%	68.62%	73.49%	71.81%	69.80%	64.84%	64.36%	67.21%	72.20%	70.33%	70.48%	64.48%	64.25%
AVG100	18.70%	32.89%	34.94%	34.71%	30.03%	32.15%	36.97%	39.75%	38.19%	37.11%	31.36%	33.93%	35.58%	38.25%	37.04%	37.33%	31.25%	33.88%

Table 18 T403 Summary of average precision across 10 query sets of entropy-weighted modes

T403	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	48.80%	78.72%	82.08%	84.41%	76.47%	78.97%	81.79%	83.83%	82.08%	83.15%	76.47%	78.55%	83.73%	85.23%	83.73%	86.01%	74.29%	78.55%
20%	42.36%	70.82%	73.71%	76.25%	67.46%	63.33%	69.88%	75.47%	74.46%	76.88%	67.27%	65.68%	71.54%	78.86%	76.01%	78.95%	69.18%	64.29%
30%	34.85%	62.15%	65.92%	60.35%	55.10%	54.87%	63.14%	69.83%	66.22%	61.90%	55.35%	54.29%	58.89%	66.87%	62.34%	68.06%	54.55%	55.23%
40%	26.40%	50.37%	53.52%	49.49%	40.95%	43.00%	54.09%	57.65%	55.17%	52.24%	41.02%	44.09%	47.57%	53.10%	50.01%	52.66%	40.79%	43.66%
50%	19.62%	35.01%	37.00%	40.61%	26.20%	35.94%	43.36%	42.73%	40.01%	42.33%	25.85%	37.34%	40.36%	41.92%	39.28%	42.33%	25.84%	37.07%
60%	16.16%	25.11%	25.80%	29.78%	18.55%	25.71%	33.82%	30.27%	30.90%	34.56%	20.03%	27.44%	32.41%	33.67%	33.56%	33.97%	19.79%	27.13%
70%	11.36%	17.20%	17.70%	20.67%	11.56%	17.94%	24.77%	22.21%	24.09%	23.79%	14.55%	21.19%	23.25%	24.96%	23.77%	23.73%	14.58%	21.17%
80%	8.36%	11.20%	10.27%	13.21%	5.33%	10.68%	17.07%	16.71%	17.79%	17.54%	9.07%	14.49%	17.20%	17.26%	17.48%	17.42%	9.13%	14.31%
90%	5.48%	5.02%	4.59%	6.62%	2.21%	5.47%	11.95%	11.01%	11.13%	12.83%	5.12%	10.67%	11.80%	13.53%	12.85%	12.77%	5.38%	10.61%
100%	1.35%	1.07%	1.24%	0.68%	0.46%	0.97%	3.06%	1.93%	2.56%	2.85%	1.52%	2.23%	2.43%	1.71%	3.26%	2.72%	1.52%	2.22%
Avg30	42.00%	70.56%	73.90%	73.67%	66.34%	65.72%	71.60%	76.38%	74.25%	73.98%	66.36%	66.17%	71.38%	76.99%	74.03%	77.67%	66.00%	66.02%
AVG100	21.35%	35.67%	37.18%	38.21%	30.43%	33.69%	40.29%	41.16%	40.44%	40.81%	31.62%	35.60%	38.92%	41.71%	40.23%	41.86%	31.50%	35.42%

Table 19 T403 Summary of average precision across 10 query sets of tfidf-weighted modes

T403	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	48.22%	78.72%	83.34%	84.41%	76.19%	79.66%	80.29%	83.83%	83.34%	83.15%	76.19%	79.66%	83.73%	84.41%	81.79%	83.15%	74.01%	79.66%
20%	42.81%	71.00%	72.47%	75.89%	68.13%	63.20%	69.70%	75.61%	73.22%	75.70%	68.23%	63.85%	69.74%	76.61%	76.50%	76.31%	70.04%	63.59%
30%	34.54%	60.80%	65.60%	59.95%	54.87%	54.39%	62.64%	68.61%	65.75%	61.60%	54.98%	54.94%	58.84%	68.22%	62.10%	65.09%	54.22%	54.46%
40%	26.19%	50.12%	54.85%	48.56%	41.31%	42.21%	52.86%	58.23%	56.92%	51.25%	41.31%	42.76%	48.92%	52.95%	50.71%	51.40%	41.11%	42.54%
50%	19.55%	35.05%	36.02%	40.27%	26.34%	37.02%	42.69%	41.56%	39.65%	42.01%	25.93%	37.26%	41.12%	42.73%	39.51%	41.92%	26.05%	37.16%
60%	16.02%	24.82%	25.80%	29.54%	18.80%	26.19%	33.77%	33.30%	31.27%	34.33%	20.56%	27.82%	32.95%	33.21%	33.05%	33.89%	20.19%	27.64%
70%	11.17%	16.88%	17.97%	20.51%	11.51%	17.88%	24.96%	24.68%	24.26%	23.79%	14.91%	20.95%	22.91%	23.21%	22.33%	23.54%	14.78%	20.98%
80%	8.28%	11.00%	10.92%	13.33%	5.29%	10.99%	16.84%	17.70%	17.77%	17.52%	9.36%	14.86%	16.77%	16.26%	16.75%	17.35%	9.27%	14.64%
90%	5.35%	4.90%	4.66%	6.41%	2.26%	5.52%	11.40%	11.56%	11.21%	12.49%	5.21%	10.69%	11.85%	12.22%	12.01%	12.38%	5.44%	10.65%
100%	1.34%	1.09%	1.23%	0.68%	0.48%	1.04%	3.18%	1.87%	2.60%	2.83%	1.56%	2.33%	2.44%	1.71%	3.62%	2.71%	1.56%	2.32%
Avg30	41.86%	70.17%	73.80%	73.42%	66.40%	65.75%	70.88%	76.02%	74.10%	73.48%	66.47%	66.15%	70.77%	76.42%	73.46%	74.85%	66.09%	65.90%
AVG100	21.21%	35.44%	37.29%	37.95%	30.52%	33.81%	39.83%	41.69%	40.60%	40.47%	31.82%	35.51%	38.93%	41.15%	39.84%	40.77%	31.67%	35.37%

Table 20 401 Average 30% APV of 54 search modes

	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Baseline	40.07%	68.92%	70.88%	69.18%	64.67%	63.76%	68.62%	73.49%	71.81%	69.80%	64.84%	64.36%	67.21%	72.20%	70.33%	70.48%	64.48%	64.25%
Entropy	42.00%	70.56%	73.90%	73.67%	66.34%	65.72%	71.60%	76.38%	74.25%	73.98%	66.36%	66.17%	71.38%	76.99%	74.03%	77.67%	66.00%	66.02%
Tf-Idf	41.86%	70.17%	73.80%	73.42%	66.40%	65.75%	70.88%	76.02%	74.10%	73.48%	66.47%	66.15%	70.77%	76.42%	73.46%	74.85%	66.09%	65.90%

Table 21 T403 Improvements

T403	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	0.00%	71.98%	76.87%	72.63%	61.38%	59.10%	71.23%	83.39%	79.19%	74.18%	61.79%	60.61%	67.71%	80.16%	75.51%	75.87%	60.91%	60.34%
entropy	4.82%	76.08%	84.42%	83.83%	65.55%	64.01%	78.68%	90.60%	85.29%	84.61%	65.60%	65.13%	78.13%	92.11%	84.73%	93.83%	64.71%	64.75%
Tf-Idf	4.45%	75.11%	84.17%	83.20%	65.69%	64.07%	76.87%	89.69%	84.91%	83.37%	65.87%	65.07%	76.60%	90.69%	83.32%	86.78%	64.93%	64.46%
AVG	3.09%	74.39%	81.82%	79.89%	64.21%	62.39%	75.59%	87.89%	83.13%	80.72%	64.42%	63.60%	74.15%	87.66%	81.18%	85.49%	63.52%	63.18%

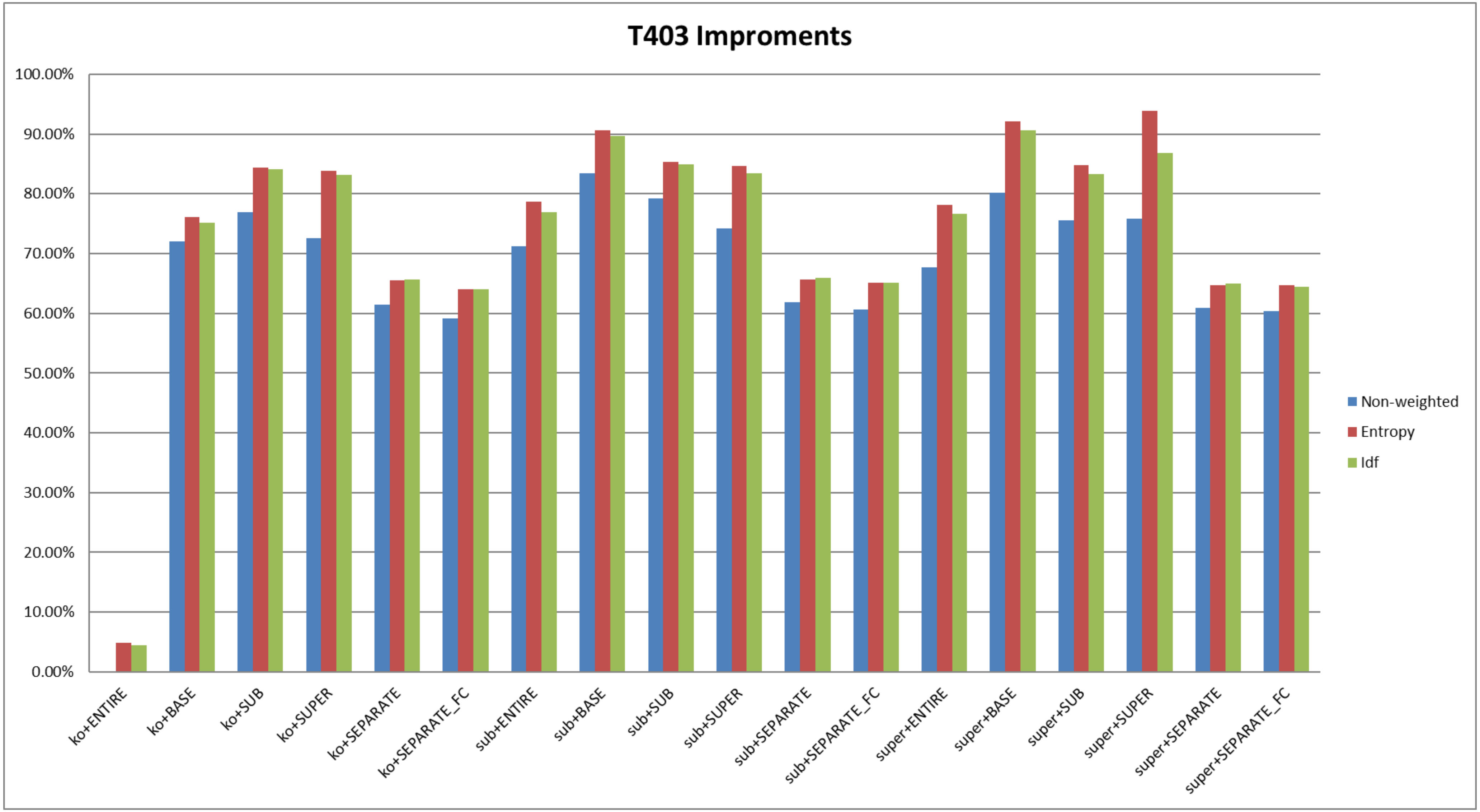


Fig. 58 T403 search mode improvements

6.2.5 Reflection on Hypotheses

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

Osteoporosis ontology was refined from three medical ontologies and this resulted in specific (less ambiguous) the concepts. The higher quality (medically-based) ontology and deeper class hierarchy offered a better query expansion potential compared with other topics. T403 experiments were successful and demonstrated remarkable OQE-based search effectiveness improvement: the AAPV (Fig 55) shows that, OQE modes provided 32% precision improvements; also, overall recall was markedly better than the baseline.

The experiment results have provided a positive indication of the search effectiveness benefits of applying OQE (SUB, SUB_SUPER) to the pure SLM. The AAPV outcomes support the hypothesis "*Topic specific small ontology context based OQE can have a positive impact on precision* ".

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

Entropy based weighting and tf-idf based weighting was applied in the search process. The results shown in subsection 6.2.3 indicate that small search effectiveness improvements were achieved.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or* *classification based local smoothing strategies are more effective than the global strategies. "*

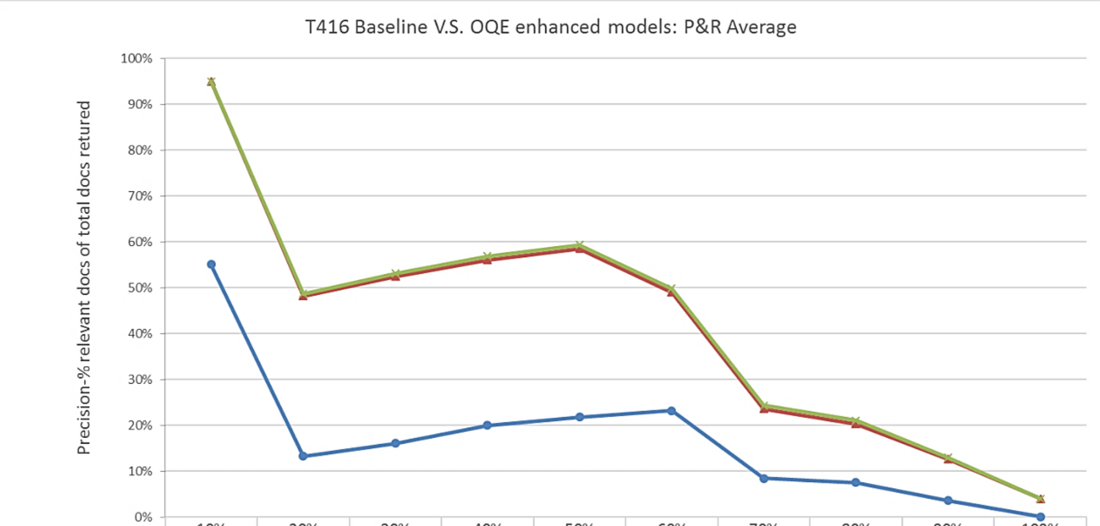
Extending osteoporosis-related terms offered an accurate document classification for the smoothing process and, as a result, provided a focused and influential use of search constraints. The benefit of document classification was shown in the experiment results, with KO+sub achieving approximately 30% improvement against the baseline. Further, OQE modes with ODC achieved even higher precision than pure OQE modes.

6.3 T416 "THREE GORGES PROJECT" EXPERIMENT RESULTS

The T416 experiments were based on 54 search modes, to provide four group comparisons in each of 10 query sets, i.e. OQE versus KO+entire, ODC versus KO+entire, KO+entire with term weighting versus KO+entire and the remaining 44 ontology enhanced modes. The combined results were considered by examining the overall average precision for all query sets (AAPV). There were 14 relevant documents to be retrieved across the entire document corpus.

6.3.1 Comparing between baseline with OQE query modes

The average results for the baseline and OQE enhanced search modes (SUB+entire, SUB_e+entire, SUPER_e+entire) are shown in Fig. 59. Dirichlet smoothing was utilised for the smoothing process in this group.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	55.14%	13.27%	16.09%	19.98%	21.83%	23.22%	8.47%	7.58%	3.64%	0.11%
sub+ENTIRE	95.00%	48.19%	52.48%	56.07%	58.50%	49.00%	23.60%	20.36%	12.67%	4.03%
super+ENTIRE	95.00%	48.74%	53.13%	56.86%	59.35%	49.89%	24.36%	21.08%	12.98%	4.03%

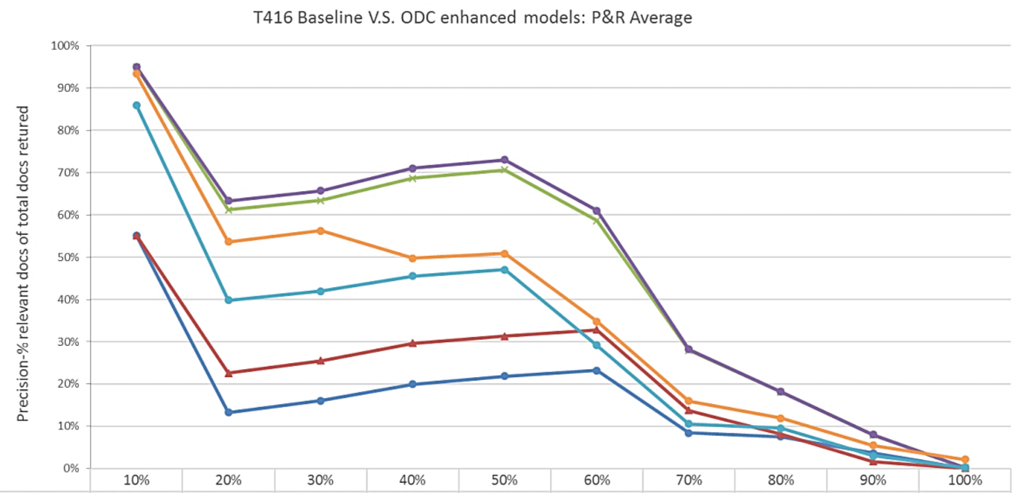
Fig. 59 T416 overall average P&R for baseline and OQE enhanced model

The precision results in Fig. 59 show OQE modes produced an excellent AAPV performance, at all recall intervals compared with baseline (28.16%). SUB+entire was 65.22% and SUPER+entire 65.62%. OQE enhanced search mode achieved about 40%

search effectiveness improvement against baseline.

6.3.2 Comparing between baseline with ODC modes

This subsection compares search effectiveness, using combined results, for the baseline and five ODC search modes. The following graphs show results based on all query sets.



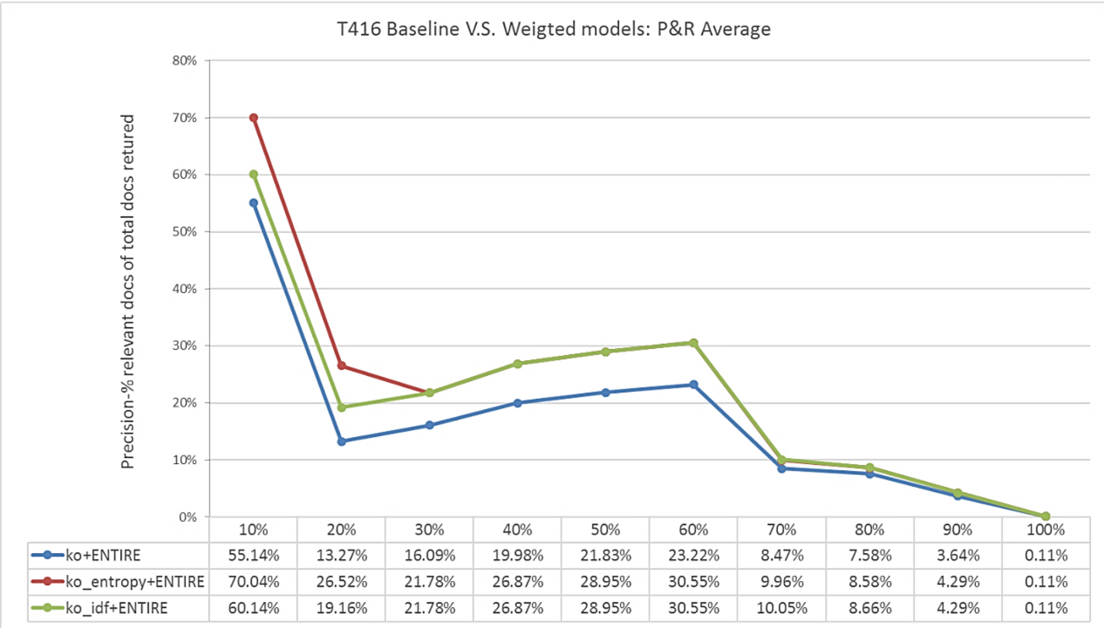
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	55.14%	13.27%	16.09%	19.98%	21.83%	23.22%	8.47%	7.58%	3.64%	0.11%
ko+BASE	55.14%	22.58%	25.47%	29.61%	31.30%	32.79%	13.72%	8.09%	1.69%	0.11%
ko+SUB	95.00%	61.18%	63.36%	68.60%	70.59%	58.59%	27.99%	18.08%	7.96%	0.26%
ko+SUPER	95.00%	63.30%	65.67%	71.05%	73.04%	61.02%	28.21%	18.18%	8.03%	0.26%
ko+SEPARATE	85.91%	39.78%	41.96%	45.53%	47.01%	29.15%	10.56%	9.57%	3.04%	0.25%
ko+SEPARATE_FC	93.33%	53.64%	56.27%	49.69%	50.86%	34.84%	15.95%	11.94%	5.52%	2.17%

Fig. 60 T416 overall average P&R for baseline and ODC enhanced model

In Fig. 60, the primary observations are that the AAPV for baseline is 28.16%, 34.40% for KO+base, 73.18% for KO+sub, 74.65% for KO+super, 55.89% for KO+separate and 67.75% for KO+separate_FC. ODC enhanced search model provide very good search performance over baseline.

6.3.3 Comparing between non-weighted with weighted modes

In this subsection, search performance is compared between the baseline and term weight-assigned baseline. Two different weighting approaches were used to generate term weighting, i.e. entropy and tf-idf. Merged P&R results of the 10 query sets are shown in Fig 61.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	55.14%	13.27%	16.09%	19.98%	21.83%	23.22%	8.47%	7.58%	3.64%	0.11%
ko_entropy+ENTIRE	70.04%	26.52%	21.78%	26.87%	28.95%	30.55%	9.96%	8.58%	4.29%	0.11%
ko_idf+ENTIRE	60.14%	19.16%	21.78%	26.87%	28.95%	30.55%	10.05%	8.66%	4.29%	0.11%

Fig. 61 T416 overall average P&R for baseline and weighting adjusted model

Fig 61 outcomes show that term weight-adjusted baseline achieved better results compared with non-weighted, with AAPV for baseline at 28.16%, KO_e+entire at 39.44% and KO_idf+entire at 33.69%. The secondary outcome was that entropy weighting was better than tf-idf weighting, but only up to 20% recall.

6.3.4 Comparison of precision results across all search modes

A comprehensive analysis of the effectiveness of 54 search modes is shown in league tables of average precision scores for each search mode (Table 22 - Table 24). The average precision values, firstly at 30% recall point and then at 100% recall, are presented in the last two rows. The primary outcome is that 53 established search modes all achieved positive results (to differing degrees) compared with baseline.

Merged AAPV values in table 25 demonstrate the search effectiveness benefit of applying ontology to SLM. SUPER_e+Super achieved the highest AAPV, achieving about 58% search effectiveness improvement (86.29% - 28.16%). Term weighting produced small improvements against the baseline.

Table 26 and Fig 62 demonstrate the relative improvement for each enhanced search mode. They indicate that the established search modes can improve search effectiveness to varying degrees. SUPER_e+Super provided highest improvement against the pure language model, i.e. 188.28%.

Table 22 T416 Summary of average precision across 10 query sets of Non-weighted modes

T416	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	55.14%	55.14%	95.00%	95.00%	85.91%	93.33%	95.00%	95.00%	95.00%	95.00%	85.91%	93.33%	95.00%	95.00%	95.00%	95.00%	85.91%	93.33%
20%	13.27%	22.58%	61.18%	63.30%	39.78%	53.64%	48.19%	60.84%	58.68%	60.80%	39.75%	53.64%	48.74%	60.20%	58.03%	60.80%	39.75%	53.64%
30%	16.09%	25.47%	63.36%	65.67%	41.96%	56.27%	52.48%	64.27%	62.40%	64.71%	41.92%	56.27%	53.13%	63.53%	61.67%	64.71%	41.92%	56.27%
40%	19.98%	29.61%	68.60%	71.05%	45.53%	49.69%	56.07%	69.10%	67.77%	70.21%	45.47%	49.69%	56.86%	68.24%	66.91%	70.21%	45.47%	49.69%
50%	21.83%	31.30%	70.59%	73.04%	47.01%	50.86%	58.50%	70.93%	69.81%	72.26%	46.95%	50.86%	59.35%	70.03%	68.91%	72.26%	46.95%	50.86%
60%	23.22%	32.79%	58.59%	61.02%	29.15%	34.84%	49.00%	60.42%	57.86%	60.29%	29.06%	34.71%	49.89%	59.50%	56.94%	60.29%	29.06%	34.71%
70%	8.47%	13.72%	27.99%	28.21%	10.56%	15.95%	23.60%	24.27%	24.72%	24.93%	10.54%	15.88%	24.36%	23.89%	24.34%	24.93%	10.45%	15.88%
80%	7.58%	8.09%	18.08%	18.18%	9.57%	11.94%	20.36%	14.93%	13.53%	13.63%	9.57%	11.96%	21.08%	15.03%	13.63%	13.63%	9.48%	11.96%
90%	3.64%	1.69%	7.96%	8.03%	3.04%	5.52%	12.67%	6.30%	6.59%	6.66%	3.04%	5.49%	12.98%	6.15%	6.45%	6.64%	2.99%	5.49%
100%	0.11%	0.11%	0.26%	0.26%	0.25%	2.17%	4.03%	0.27%	0.17%	0.17%	0.25%	2.16%	4.03%	0.27%	0.17%	0.17%	0.25%	2.16%
Avg30	28.16%	34.40%	73.18%	74.65%	55.89%	67.75%	65.22%	73.37%	72.03%	73.50%	55.86%	67.75%	65.62%	72.91%	71.57%	73.50%	55.86%	67.75%
AVG100	16.96%	22.05%	47.16%	48.38%	31.28%	37.42%	41.99%	46.63%	45.65%	46.87%	31.25%	37.40%	42.54%	46.18%	45.20%	46.86%	31.22%	37.40%

Table 23 T416 Summary of average precision across 10 query sets of Entropy-weighted modes

T416	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	70.04%	60.14%	100.00%	100.00%	91.43%	95.00%	100.00%	95.00%	100.00%	100.00%	91.43%	95.00%	100.00%	95.00%	100.00%	100.00%	91.43%	95.00%
20%	26.52%	31.16%	69.89%	73.21%	37.66%	58.65%	61.18%	72.39%	68.39%	71.71%	40.13%	58.64%	61.75%	73.89%	73.48%	78.21%	40.13%	58.64%
30%	21.78%	34.08%	72.08%	75.33%	41.28%	60.86%	65.35%	75.41%	72.08%	75.33%	43.23%	60.86%	65.90%	76.75%	76.30%	80.67%	43.23%	60.86%
40%	26.87%	38.10%	77.17%	80.17%	47.92%	53.52%	67.95%	79.67%	77.17%	80.17%	47.86%	53.52%	68.42%	80.74%	77.79%	80.17%	47.86%	53.52%
50%	28.95%	39.68%	79.03%	81.89%	49.72%	54.38%	70.20%	81.25%	79.03%	81.89%	49.66%	54.38%	70.63%	82.22%	79.55%	81.89%	49.66%	54.38%
60%	30.55%	41.03%	60.80%	63.54%	29.46%	40.95%	60.10%	60.52%	60.75%	63.49%	29.38%	40.68%	60.49%	61.41%	61.20%	63.49%	29.38%	40.68%
70%	9.96%	15.12%	30.19%	31.45%	8.26%	12.76%	24.77%	27.30%	27.78%	29.04%	8.24%	12.70%	24.89%	28.56%	29.04%	29.04%	8.22%	12.70%
80%	8.58%	5.46%	19.12%	19.30%	8.29%	8.21%	21.28%	13.97%	13.39%	13.58%	8.29%	8.25%	21.64%	13.80%	13.22%	13.58%	8.26%	8.25%
90%	4.29%	1.80%	8.79%	9.03%	2.49%	5.50%	12.08%	6.74%	7.06%	7.30%	2.49%	5.46%	12.37%	6.71%	7.04%	7.30%	2.49%	5.46%
100%	0.11%	0.11%	0.31%	0.31%	0.25%	2.22%	4.66%	0.28%	0.18%	0.17%	0.25%	2.21%	4.66%	0.28%	0.18%	0.17%	0.25%	2.21%
Avg30	39.44%	41.79%	80.66%	82.85%	56.79%	71.50%	75.51%	80.94%	80.16%	82.35%	58.26%	71.50%	75.88%	81.88%	83.26%	86.29%	58.26%	71.50%
AVG100	22.79%	26.67%	51.74%	53.42%	31.68%	39.20%	48.76%	51.25%	50.58%	52.27%	32.10%	39.17%	49.07%	51.94%	51.78%	53.45%	32.09%	39.17%

Table 24 T416 Summary of average precision across 10 query sets of idf-weighted modes

T416	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	60.14%	60.14%	100.00%	100.00%	91.25%	95.00%	100.00%	95.00%	100.00%	100.00%	91.25%	95.00%	100.00%	95.00%	100.00%	100.00%	91.25%	95.00%
20%	19.16%	31.16%	68.89%	71.73%	39.83%	58.64%	61.04%	71.39%	67.39%	70.23%	39.79%	58.64%	61.61%	72.98%	68.98%	74.23%	39.79%	58.64%
30%	21.78%	34.08%	71.13%	73.78%	42.91%	60.86%	65.18%	74.46%	71.13%	73.78%	42.87%	60.86%	65.73%	75.74%	72.41%	77.11%	42.87%	60.86%
40%	26.87%	38.10%	76.34%	78.63%	47.53%	53.52%	67.74%	78.84%	76.34%	78.63%	47.47%	53.52%	68.21%	79.67%	77.17%	78.63%	47.47%	53.52%
50%	28.95%	39.68%	78.25%	80.38%	49.33%	54.38%	69.97%	80.47%	78.25%	80.38%	49.27%	54.38%	70.40%	81.15%	78.93%	80.38%	49.27%	54.38%
60%	30.55%	41.03%	60.48%	62.48%	28.70%	40.77%	59.95%	59.86%	60.42%	62.41%	28.54%	40.53%	60.35%	60.41%	60.96%	62.41%	28.54%	40.53%
70%	10.05%	15.48%	26.05%	31.19%	8.31%	18.09%	24.75%	23.16%	24.91%	30.05%	8.29%	18.02%	24.63%	28.31%	30.05%	30.05%	8.26%	18.02%
80%	8.66%	6.66%	14.92%	19.30%	8.33%	13.75%	21.28%	10.67%	10.15%	14.53%	8.33%	13.78%	21.54%	14.70%	14.18%	14.53%	8.30%	13.78%
90%	4.29%	2.09%	7.09%	9.16%	2.49%	5.48%	12.07%	5.03%	5.33%	7.40%	2.50%	5.43%	12.28%	6.83%	7.14%	7.40%	2.49%	5.43%
100%	0.11%	0.11%	0.31%	0.31%	0.25%	2.20%	4.65%	0.28%	0.18%	0.18%	0.25%	2.20%	4.65%	0.28%	0.18%	0.17%	0.25%	2.20%
Avg30	33.69%	41.79%	80.01%	81.83%	58.00%	71.50%	75.41%	80.28%	79.51%	81.33%	57.97%	71.50%	75.78%	81.24%	80.46%	83.78%	57.97%	71.50%
AVG100	21.08%	26.85%	50.35%	52.70%	31.89%	40.27%	48.66%	49.92%	49.41%	51.76%	31.86%	40.24%	48.94%	51.51%	51.00%	52.49%	31.85%	40.24%

Table 25 T416 Average 30% APV of 54 search modes

	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Baseline	28.16%	34.40%	73.18%	74.65%	55.89%	67.75%	65.22%	73.37%	72.03%	73.50%	55.86%	67.75%	65.62%	72.91%	71.57%	73.50%	55.86%	67.75%
Entropy	39.44%	41.79%	80.66%	82.85%	56.79%	71.50%	75.51%	80.94%	80.16%	82.35%	58.26%	71.50%	75.88%	81.88%	83.26%	86.29%	58.26%	71.50%
Tf-Idf	33.69%	41.79%	80.01%	81.83%	58.00%	71.50%	75.41%	80.28%	79.51%	81.33%	57.97%	71.50%	75.78%	81.24%	80.46%	83.78%	57.97%	71.50%

Table 26 T416 improvment

T416	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	0.00%	22.13%	159.82%	165.06%	98.42%	140.54%	131.58%	160.50%	155.73%	160.98%	98.34%	140.54%	133.00%	158.86%	154.10%	160.98%	98.34%	140.54%
entropy	40.05%	48.39%	186.38%	194.16%	101.63%	153.87%	168.10%	187.37%	184.60%	192.38%	106.87%	153.87%	169.42%	190.72%	195.62%	206.39%	106.87%	153.87%
Tf-Idf	19.62%	48.39%	184.07%	190.56%	105.92%	153.87%	167.74%	185.06%	182.29%	188.78%	105.83%	153.86%	169.07%	188.45%	185.68%	197.46%	105.83%	153.86%
AVG	19.89%	39.64%	176.76%	183.26%	101.99%	149.42%	155.81%	177.64%	174.21%	180.71%	103.68%	149.42%	157.16%	179.34%	178.47%	188.28%	103.68%	149.42%

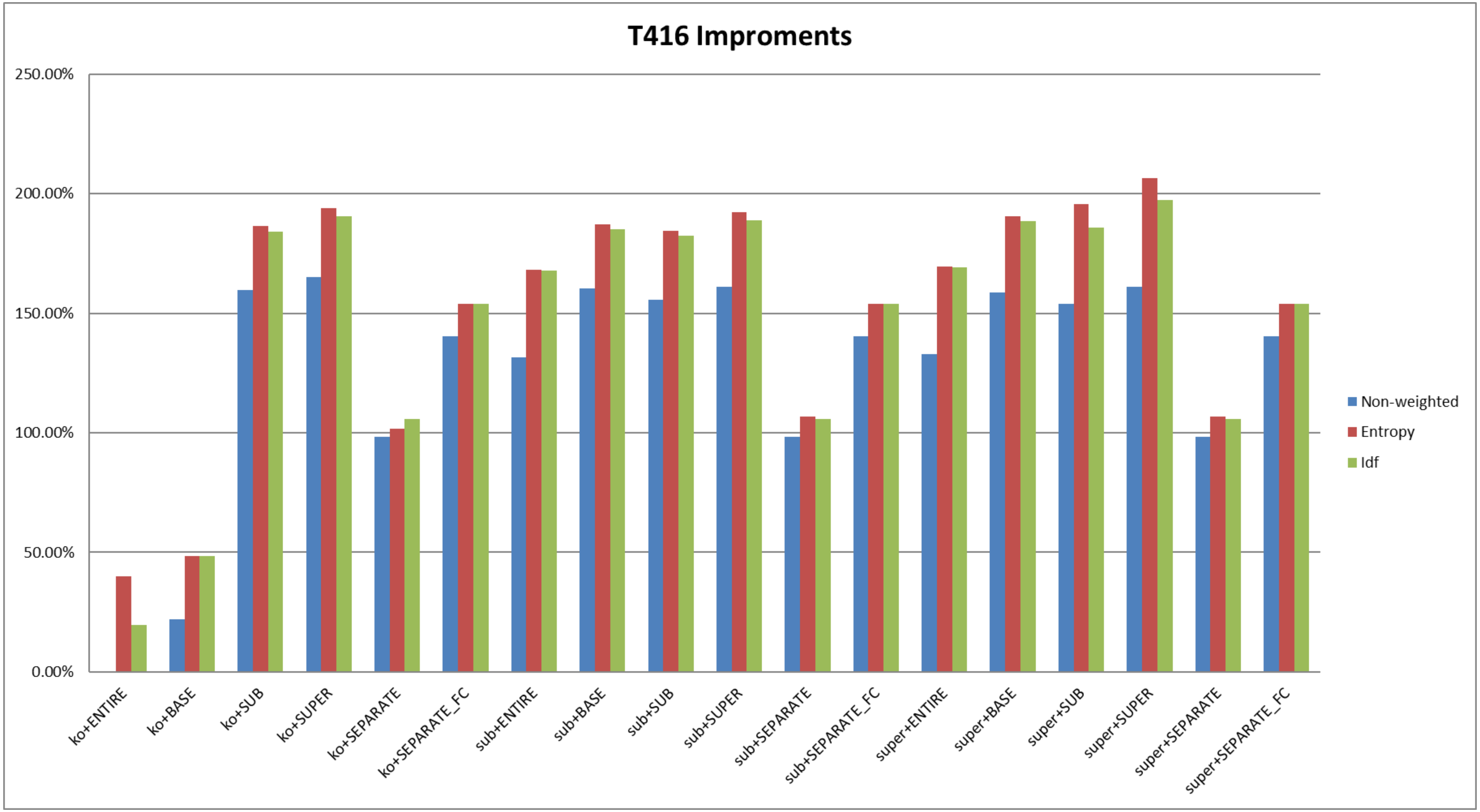


Fig. 62 T416 search mode improvements

6.3.5 Reflection on Hypotheses

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

The Hydro-electric ontology was not developed specifically for the "Three gorges project". T416 experiment results in subsection 6.3.1 show that the OQE enhanced search model produced higher AAPV against keyword-only search. The AAPV of SUPER+entire was approximately 37% higher than baseline. Whilst, the analysis of individual query results in appendix C, shows overall recalls were better than keyword only modes, with 2 of 10 query sets achieving better recall.

These outcomes demonstrate the benefit of applying the query expansion beyond baseline, particularly using SUPER+entire, and provide clear evidence to support the hypothesis, in terms of both precision and recall.

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

As in the previous two topic experiments, entropy and tf-idf algorithms employed term weighting to reduce the effect of expanded, ambiguous terms. The results in subsection 6.3.3 showed that assigning lower weightings to expanded terms can enhance retrieval performance to some extent. It is considered that the hypothesis has been supported.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document* *clustering or classification based local smoothing strategies are more effective than the global strategies. "*

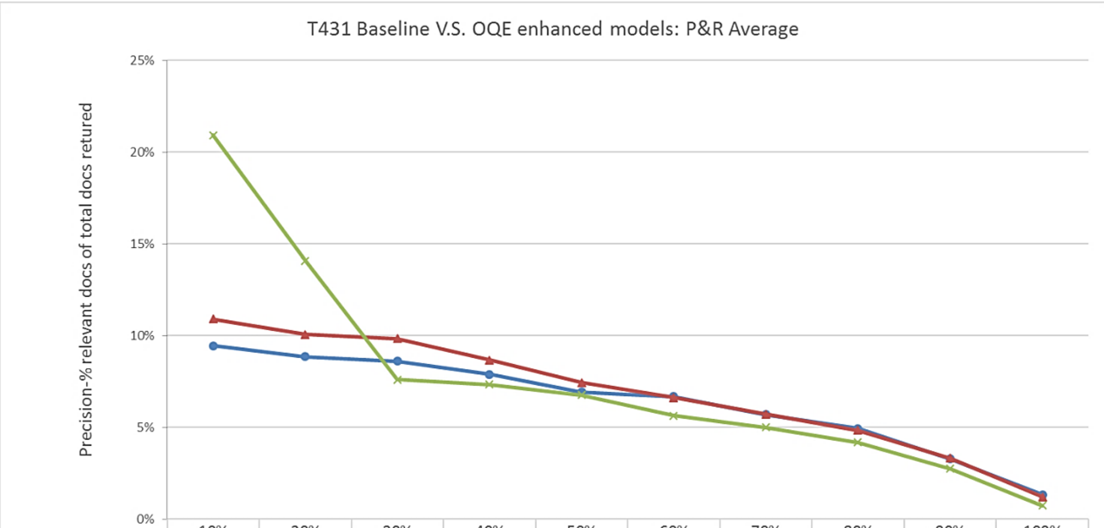
The ontology based document classification (Base, Sub, Super, Separate and Separate_FC) provided the corpus structure for local smoothing. The benefit of extending document classification was indicated through the experiment results, i.e. KO+separate achieved about 46% improvement against the baseline. As Fig 60 showed in subsection 6.3.2, all ODC enhanced search modes demonstrate signification improvement over keyword only search mode. Overall, the results provide good evidence to support the hypothesis.

6.4 T431 "ROBOTIC TECHNOLOGY" EXPERIMENT RESULTS

The following T431 experiment P&R graphs demonstrate the search effectiveness of 54 search modes. As with previous topics, results comparisons are divided into four groups, i.e. OQE, term weighting and ODC enhance SLM are compared with the pure SLM (KO+entire). There are 148 relevant T431 documents across 247,491 documents. The graph y-axes ranges have been reduced for presentation purposes, as many maximum precision values were below 30%.

6.4.1 Comparing between baseline with OQE query modes

This subsection compares the merged experiment results for the baseline and two OQE search modes. The P&R graph comparing the combined results for 3 modes across the 10 query sets are shown in Fig. 63.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	9.44%	8.85%	8.61%	7.88%	6.93%	6.69%	5.70%	4.94%	3.29%	1.33%
sub+ENTIRE	10.90%	10.07%	9.84%	8.67%	7.43%	6.62%	5.72%	4.85%	3.33%	1.21%
super+ENTIRE	20.93%	14.08%	7.60%	7.35%	6.76%	5.64%	5.01%	4.19%	2.75%	0.74%

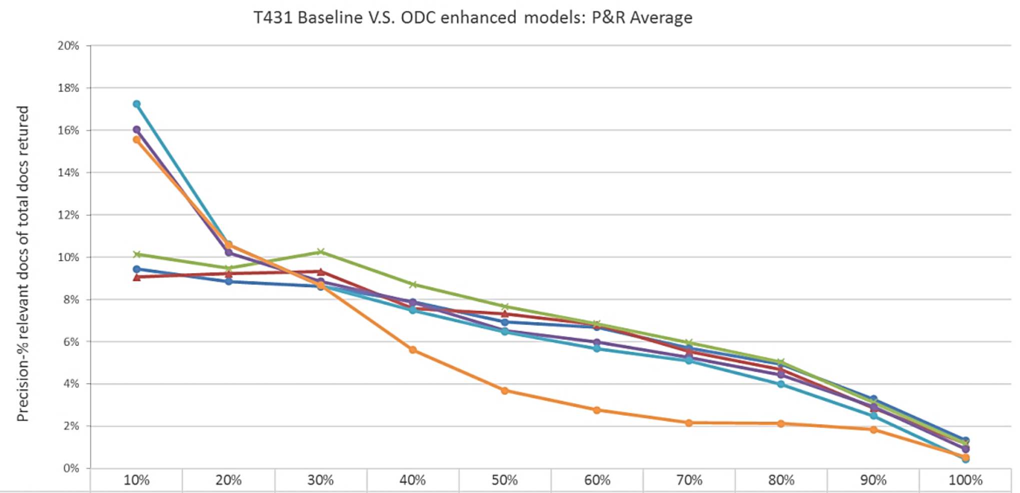
Fig. 63 T431 overall average P&R for baseline and OQE enhanced model

Based on the average number of the relevant documents returned, the primary outcomes were that the AAPV for KO+ Entire (baseline) was 8.97%, with 10.27% for SUB+entire, and 14.21% for SUPER+entire. The AAPVs indicate that SUB_SUPER

OQE mode improvement was markedly better, up to the 20% recall level.

6.4.2 Comparing between baseline with ODC modes

The following graphs show combined P&R results for comparison between baseline and five ODC modes.



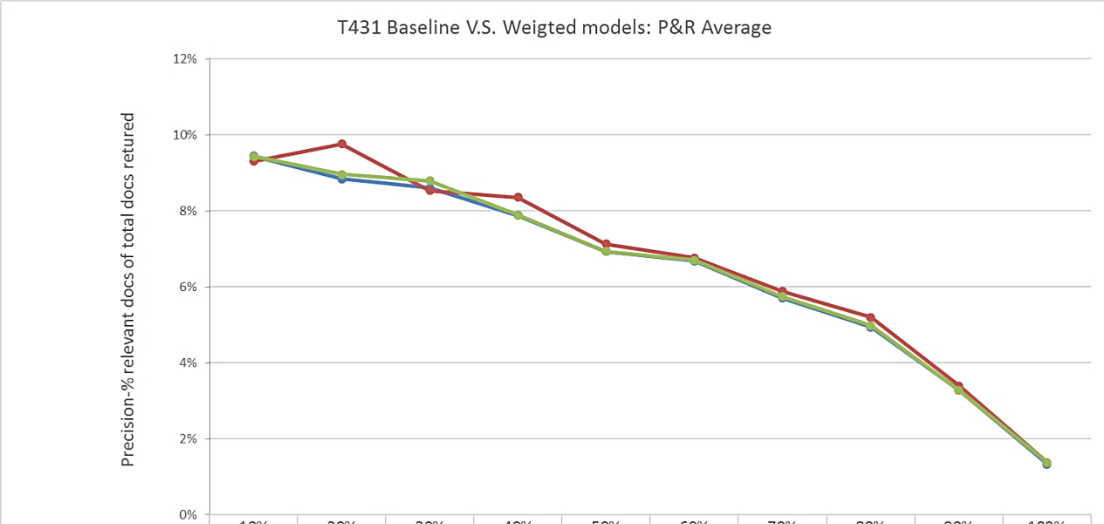
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	9.44%	8.85%	8.61%	7.88%	6.93%	6.69%	5.70%	4.94%	3.29%	1.33%
ko+BASE	9.05%	9.22%	9.32%	7.58%	7.33%	6.80%	5.53%	4.68%	2.86%	1.18%
ko+SUB	10.14%	9.47%	10.25%	8.71%	7.65%	6.86%	5.96%	5.03%	3.13%	1.20%
ko+SUPER	16.03%	10.22%	8.86%	7.85%	6.52%	5.98%	5.26%	4.43%	2.92%	0.93%
ko+SEPARATE	17.24%	10.59%	8.65%	7.49%	6.47%	5.68%	5.09%	3.99%	2.50%	0.45%
ko+SEPARATE_FC	15.55%	10.59%	8.65%	5.62%	3.69%	2.77%	2.17%	2.13%	1.85%	0.54%

Fig. 64 T431 overall average P&R for baseline and ODC enhanced models

Fig. 64 shows the results of the combined 10 queries for ODC. It shows that search effectiveness improvement was achieved, with 8.97% for KO+entire, 9.20% for KO+base, 9.96% for KO+sub, 11.70% for KO+SUPER, 12.16% for KO+separate and 11.60% for KO+separate_FC. The secondary outcome was that KO+separate_FC mode produced worse results between 30% and 100% recall intervals; this reflects an inaccurate document classification, by using an ontology-based separate document classification with term frequency.

6.4.3 Comparing between non-weighted with weighted modes

This subsection will assess the search effectiveness of applying term weighting to keywords. 3 search mode (baseline, KO_e+entire, KO_idf+entire) results are illustrated in Fig 65. It should be pointed out that, the smoothing approach used in the three search modes was Dirichlet smoothing.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	9.44%	8.85%	8.61%	7.88%	6.93%	6.69%	5.70%	4.94%	3.29%	1.33%
ko_entropy+ENTIRE	9.31%	9.76%	8.54%	8.36%	7.13%	6.76%	5.88%	5.19%	3.39%	1.37%
ko_idf+ENTIRE	9.43%	8.96%	8.80%	7.89%	6.93%	6.70%	5.74%	4.98%	3.28%	1.37%

Fig. 65 T431 overall average P&R for baseline and weighting adjusted models

Search results show that weight-assigned search modes achieved little improvement against the baseline, i.e. AAPV for baseline was 8.97%, with 9.20% for entropy adjusted mode and 9.06% for tf-idf mode.

6.4.4 Comparison of precision results across all search modes

A league table of average precision score was developed to assess the search effectiveness, for each of 54 search modes, by comparing all average recall points. As with previous experiments, gradient colours from red to green show the descending of average precision values. Again, the average precision value of the first 30% and 100% recall points of each search mode are presented in the last two rows of Table 27, Table 28 and Table 29. The primary outcome was that extending OQE, term weighting, and ODC to SLM produced significant search effectiveness improvements.

Merged AAPV results in Table 30 indicate that OQE and ontology-based document classification enhanced smoothing can have a positive impact on pure SLM searching. The best search model in T416 was SUPER_e+Super, which achieved approximately 7% search effectiveness improvement beyond the baseline. A secondary outcome from the overall query set results was that OQE can produce better recall, as demonstrated in individual query search results in Appendix C.

In Fig 66, the primary observation is that every enhanced mode improved the search effectiveness - to a varying degree. SUPER_e+Super provided 70.60% improvement against baseline, which was the best result from all 53 established modes.

Table 27 T431 Summary of average precision across 10 query sets of Non-weighted modes

T431	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	9.44%	9.05%	10.14%	16.03%	17.24%	15.55%	10.90%	13.26%	10.39%	16.33%	16.71%	16.02%	20.93%	18.69%	21.70%	21.75%	19.75%	18.59%
20%	8.85%	9.22%	9.47%	10.22%	10.59%	10.59%	10.07%	8.51%	9.68%	12.10%	11.46%	11.82%	14.08%	13.65%	14.73%	14.52%	15.03%	14.62%
30%	8.61%	9.32%	10.25%	8.86%	8.65%	8.65%	9.84%	8.52%	10.39%	9.05%	9.34%	8.67%	7.60%	9.03%	8.76%	9.13%	10.13%	8.47%
40%	7.88%	7.58%	8.71%	7.85%	7.49%	5.62%	8.67%	7.71%	8.79%	7.93%	7.78%	5.66%	7.35%	7.61%	7.52%	7.59%	8.84%	5.24%
50%	6.93%	7.33%	7.65%	6.52%	6.47%	3.69%	7.43%	6.92%	7.68%	6.55%	6.63%	3.75%	6.76%	6.62%	6.47%	6.65%	7.53%	4.20%
60%	6.69%	6.80%	6.86%	5.98%	5.68%	2.77%	6.62%	6.19%	6.86%	5.98%	5.77%	2.83%	5.64%	5.79%	5.90%	5.90%	6.46%	3.20%
70%	5.70%	5.53%	5.96%	5.26%	5.09%	2.17%	5.72%	5.29%	5.95%	5.27%	5.15%	2.21%	5.01%	5.35%	5.29%	5.30%	5.86%	2.91%
80%	4.94%	4.68%	5.03%	4.43%	3.99%	2.13%	4.85%	4.62%	4.98%	4.40%	4.00%	2.14%	4.19%	4.49%	4.34%	4.32%	4.70%	2.65%
90%	3.29%	2.86%	3.13%	2.92%	2.50%	1.85%	3.33%	2.72%	3.08%	2.88%	2.50%	1.82%	2.75%	2.99%	2.71%	2.71%	2.69%	1.89%
100%	1.33%	1.18%	1.20%	0.93%	0.45%	0.54%	1.21%	1.19%	1.06%	0.83%	0.38%	0.51%	0.74%	1.19%	0.64%	0.64%	0.63%	0.54%
Avg30	8.97%	9.20%	9.96%	11.70%	12.16%	11.60%	10.27%	10.10%	10.16%	12.50%	12.50%	12.17%	14.21%	13.79%	15.06%	15.13%	14.97%	13.89%
AVG100	6.37%	6.35%	6.84%	6.90%	6.82%	5.36%	6.87%	6.49%	6.89%	7.13%	6.97%	5.54%	7.50%	7.54%	7.81%	7.85%	8.16%	6.23%

Table 28 T431 Summary of average precision across 10 query sets of Entropy-weighted modes

T431	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	9.31%	10.04%	11.20%	15.68%	18.61%	15.84%	12.53%	14.26%	12.84%	16.23%	18.20%	17.25%	21.15%	18.43%	22.58%	22.56%	21.21%	19.56%
20%	9.76%	9.56%	9.77%	10.30%	11.44%	11.15%	10.21%	9.25%	9.67%	11.81%	12.06%	11.45%	13.90%	13.74%	14.17%	14.18%	14.95%	14.04%
30%	8.54%	9.47%	10.20%	9.45%	8.84%	8.52%	8.75%	8.32%	9.26%	9.58%	9.36%	8.66%	9.75%	9.64%	9.59%	9.84%	10.02%	8.36%
40%	8.36%	8.50%	9.34%	8.86%	8.30%	6.24%	7.93%	7.61%	8.62%	8.97%	8.61%	6.28%	7.72%	8.87%	8.64%	8.66%	9.40%	5.83%
50%	7.13%	7.49%	7.99%	7.43%	7.08%	4.16%	6.72%	6.36%	6.99%	7.47%	7.24%	4.20%	6.98%	7.82%	7.55%	7.56%	8.26%	4.93%
60%	6.76%	6.99%	6.91%	6.11%	6.05%	2.27%	5.78%	5.54%	5.99%	6.08%	6.15%	2.33%	5.95%	6.51%	6.18%	6.19%	6.83%	3.48%
70%	5.88%	6.09%	6.23%	5.59%	5.31%	2.11%	5.17%	4.85%	5.47%	5.55%	5.38%	2.13%	5.48%	5.53%	5.70%	5.68%	6.14%	3.08%
80%	5.19%	4.88%	5.06%	4.58%	4.00%	1.90%	4.23%	3.92%	4.40%	4.49%	4.01%	1.88%	4.10%	4.41%	4.36%	4.32%	4.69%	2.54%
90%	3.39%	3.15%	3.13%	2.96%	2.27%	1.63%	2.73%	2.73%	2.76%	2.88%	2.27%	1.56%	2.79%	2.92%	2.88%	2.84%	2.49%	1.92%
100%	1.37%	1.18%	1.20%	0.88%	0.66%	0.71%	0.76%	1.21%	0.77%	0.77%	0.54%	0.63%	0.75%	1.26%	0.71%	0.71%	0.80%	0.45%
Avg30	9.20%	9.69%	10.39%	11.81%	12.96%	11.84%	10.50%	10.61%	10.59%	12.54%	13.21%	12.45%	14.93%	13.94%	15.45%	15.53%	15.39%	13.98%
AVG100	6.55%	6.73%	7.10%	7.18%	7.26%	5.45%	6.48%	6.40%	6.68%	7.38%	7.38%	5.64%	7.86%	7.91%	8.24%	8.25%	8.48%	6.42%

Table 29 T431 Summary of average precision across 10 query sets of tfidf-weighted modes

T431	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	9.43%	9.05%	10.17%	16.40%	17.45%	15.69%	10.99%	13.43%	10.55%	16.52%	17.22%	16.01%	18.93%	18.69%	21.96%	21.93%	19.38%	18.42%
20%	8.96%	9.29%	9.61%	10.29%	10.80%	10.58%	10.06%	8.59%	9.85%	12.18%	11.43%	11.74%	14.03%	13.81%	14.73%	14.57%	14.75%	14.54%
30%	8.80%	9.59%	10.15%	8.71%	8.72%	8.63%	9.79%	8.40%	10.20%	8.89%	9.28%	8.79%	9.09%	8.90%	8.75%	9.19%	10.34%	8.55%
40%	7.89%	7.68%	8.63%	7.93%	7.73%	5.77%	8.97%	7.55%	8.74%	8.01%	8.01%	5.81%	8.08%	7.42%	7.50%	7.59%	9.09%	5.51%
50%	6.93%	7.40%	7.78%	6.78%	6.73%	3.97%	7.50%	6.38%	7.81%	6.81%	6.90%	4.28%	7.04%	6.34%	6.51%	6.66%	7.82%	4.23%
60%	6.70%	6.95%	6.84%	5.97%	5.97%	2.91%	6.60%	5.61%	6.86%	5.99%	6.06%	2.95%	6.81%	5.71%	5.89%	5.90%	6.72%	3.20%
70%	5.74%	5.62%	5.84%	5.19%	5.22%	2.16%	5.70%	4.96%	5.84%	5.20%	5.28%	2.21%	5.83%	5.34%	5.29%	5.29%	6.06%	2.95%
80%	4.98%	4.81%	5.02%	4.49%	4.17%	2.12%	4.87%	4.43%	4.97%	4.45%	4.18%	2.13%	5.00%	4.49%	4.32%	4.31%	4.91%	2.65%
90%	3.28%	3.05%	3.36%	2.99%	2.66%	1.85%	3.34%	2.80%	3.24%	2.95%	2.66%	1.82%	3.28%	2.87%	2.69%	2.70%	2.86%	1.88%
100%	1.37%	1.18%	1.22%	0.91%	0.46%	0.60%	1.20%	1.19%	1.09%	0.81%	0.39%	0.56%	1.37%	1.17%	0.65%	0.65%	0.64%	0.54%
Avg30	9.06%	9.31%	9.98%	11.80%	12.32%	11.63%	10.28%	10.14%	10.20%	12.53%	12.64%	12.18%	14.02%	13.80%	15.15%	15.23%	14.83%	13.84%
AVG100	6.41%	6.46%	6.86%	6.97%	6.99%	5.43%	6.90%	6.34%	6.91%	7.18%	7.14%	5.63%	7.94%	7.47%	7.83%	7.88%	8.26%	6.25%

Table 30 T431 Average 30% APV of 54 search modes

	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	8.97%	9.20%	9.96%	11.70%	12.16%	11.60%	10.27%	10.10%	10.16%	12.50%	12.50%	12.17%	14.21%	13.79%	15.06%	15.13%	14.97%	13.89%
entropy	9.20%	9.69%	10.39%	11.81%	12.96%	11.84%	10.50%	10.61%	10.59%	12.54%	13.21%	12.45%	14.93%	13.94%	15.45%	15.53%	15.39%	13.98%
Tf-Idf	9.06%	9.31%	9.98%	11.80%	12.32%	11.63%	10.28%	10.14%	10.20%	12.53%	12.64%	12.18%	14.02%	13.80%	15.15%	15.23%	14.83%	13.84%

Table 31 T431 improvments

T431	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	0.00%	2.56%	11.03%	30.51%	35.64%	29.35%	14.57%	12.60%	13.26%	39.38%	39.44%	35.78%	58.43%	53.82%	67.99%	68.78%	66.97%	54.93%
entropy	2.65%	8.07%	15.87%	31.73%	44.59%	32.02%	17.10%	18.32%	18.12%	39.85%	47.31%	38.90%	66.51%	55.42%	72.27%	73.17%	71.68%	55.97%
Tf-Idf	1.07%	3.85%	11.27%	31.61%	37.44%	29.75%	14.65%	13.11%	13.77%	39.76%	41.02%	35.82%	56.32%	53.92%	68.93%	69.86%	65.34%	54.34%
AVG	1.24%	4.83%	12.72%	31.29%	39.22%	30.37%	15.44%	14.67%	15.05%	39.66%	42.59%	36.83%	60.42%	54.39%	69.73%	70.60%	68.00%	55.08%

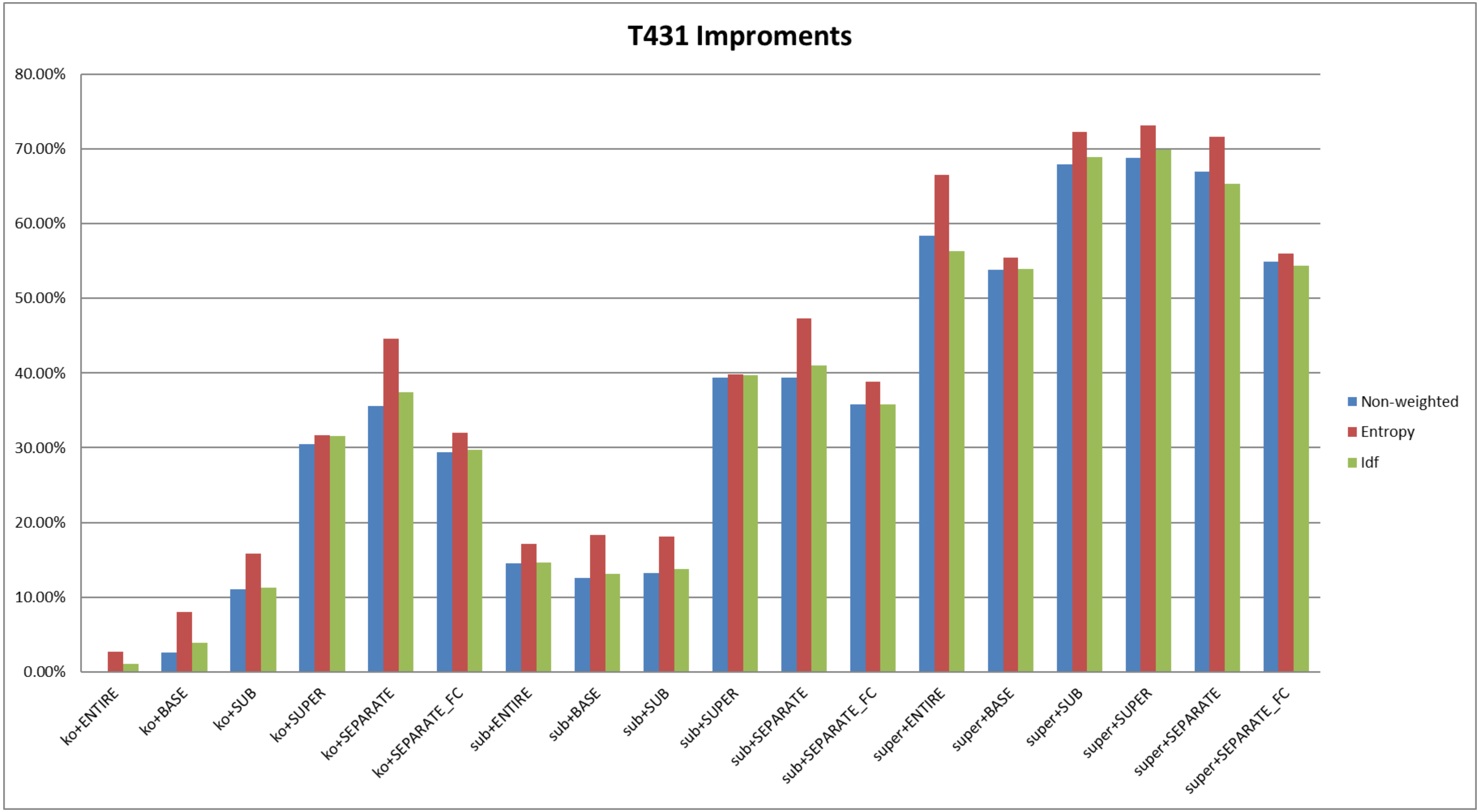


Fig. 66 T431 search mode improvements

6.4.5 Reflection on Hypotheses

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

T431's ontology was not specifically designed for the topic, although it is related to robotics and contains robotic domain related classes. However, a relevant document was considered as one identified with robotic applications not robotic technology.

The T431 experiments were successful and indicate that OQE-based search can provide better search effectiveness against keyword-only, i.e. all OQE-based search modes improved search effectiveness to varying degrees (subsection 6.4.1). The experiment results have demonstrated the benefit of applying OQE to SLM. The experiment results fully support the hypothesis.

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

Two weighting approaches were used to eliminate ambiguous expanded terms, the experiment results in subsection 6.4.3 show that using term weighting can improve the search performance, to some extent, especially with entropy weighting, where it markedly improved search effectiveness.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or classification based local smoothing strategies are more effective than the global strategies. "*

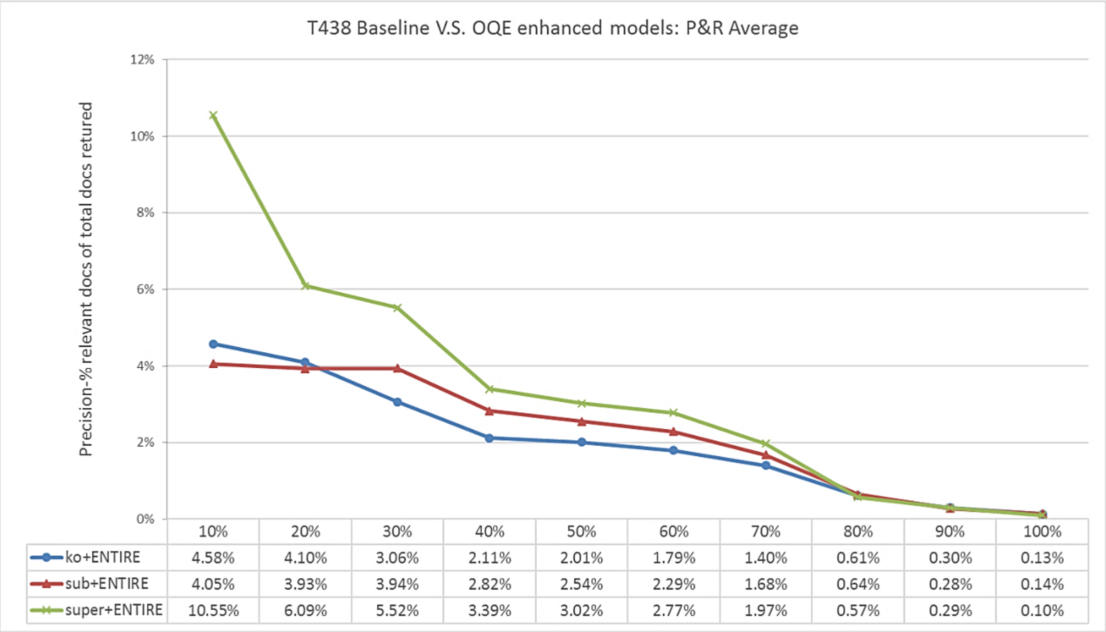
The experiment primarily shows that positive results were achieved from extending ontology-based document classification strategy to SLM. T431's ODC AAPV results demonstrated remarkable search effectiveness improvements over baseline. Nevertheless, ODC combined with SUPER OQE achieved the best results, with approximately 70% improvement over baseline (Fig 66). Overall, the experiment results support hypothesis (iii).

6.5 T438 "TOURISM, INCREASE" EXPERIMENT RESULTS

This section presents the T438results, where 46 relevant documents were targeted across the document collection. The following P&R graphs review 54 search modes in four groups, as in previous 4 topics. The results evaluate the overall group of 10 query sets. It should be noted that, because of very low precision levels, the y-axes were reduced for better visual understanding.

6.5.1 Comparing between baseline with OQE query modes

The following T438 P&R graph reviews were based on 3 search modes to provide comparisons between the baseline and two OQE modes, i.e. SUB+entire, SUPER+entire. The Dirichlet smoothing approach was applied in this group, to eliminate the data sparseness. Fig. 67 shows the P&R graph comparing the combined results for 3 modes across 10 query sets.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	4.58%	4.10%	3.06%	2.11%	2.01%	1.79%	1.40%	0.61%	0.30%	0.13%
sub+ENTIRE	4.05%	3.93%	3.94%	2.82%	2.54%	2.29%	1.68%	0.64%	0.28%	0.14%
super+ENTIRE	10.55%	6.09%	5.52%	3.39%	3.02%	2.77%	1.97%	0.57%	0.29%	0.10%

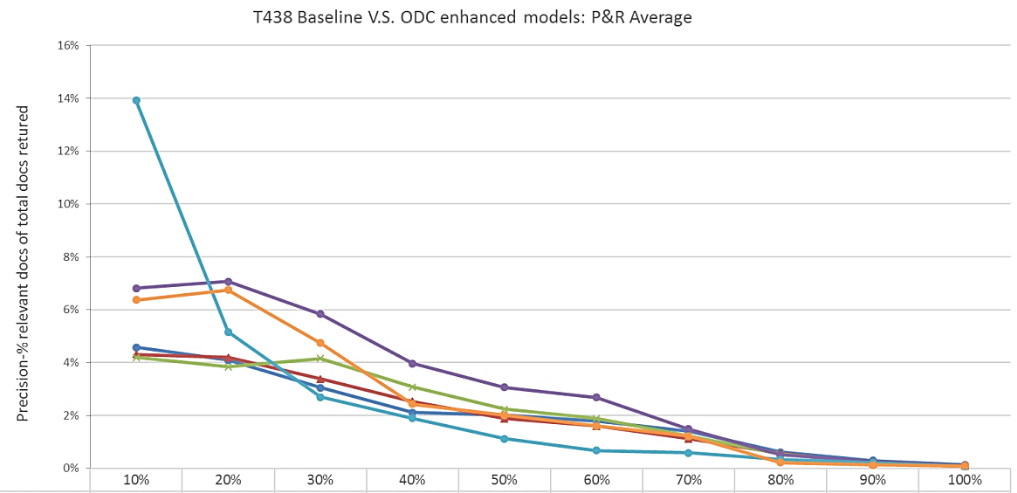
Fig. 67 T438 overall average P&R for baseline and OQE enhanced model

Based on the average precision of each recall point, the primary outcome was that the AAPV for KO+entire was 3.91%, with 3.97% for SUB+entire and 7.39% for

SUPER+entire. The SUB_SUPER OQE modes show a positive AAPV outcome compared to pure SLM. The secondary outcome was that subclass-based OQE produced worse results in the first 20% recall intervals; this might have been caused by ambiguous expanded terms from the tourism ontology.

6.5.2 Comparing between baseline with ODC modes

The following P&R graphs compare combined results, to evaluate ODC mode effectiveness. The KO+entire is compared with five ODC modes, i.e. KO+base, KO+sub, KO+super, KO+separate and KO+separate_FC. The overall AAPV of all 10 query sets are shown in Fig. 68.



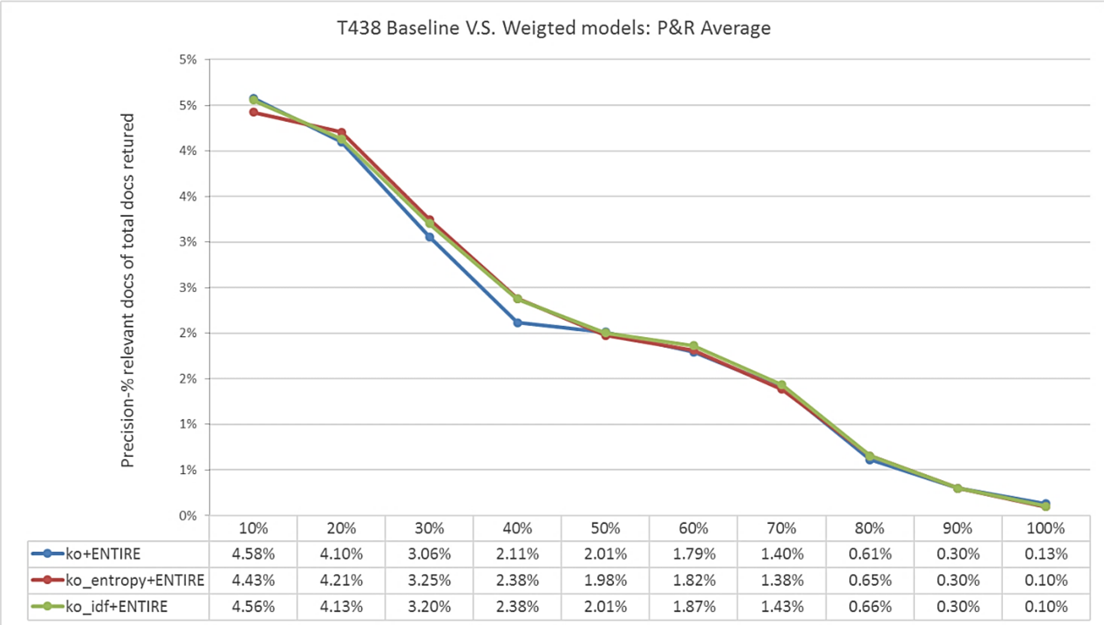
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	4.58%	4.10%	3.06%	2.11%	2.01%	1.79%	1.40%	0.61%	0.30%	0.13%
ko+BASE	4.31%	4.19%	3.38%	2.53%	1.88%	1.61%	1.12%	0.57%	0.18%	0.09%
ko+SUB	4.19%	3.84%	4.16%	3.08%	2.24%	1.88%	1.22%	0.54%	0.21%	0.08%
ko+SUPER	6.81%	7.07%	5.83%	3.97%	3.06%	2.68%	1.49%	0.52%	0.25%	0.12%
ko+SEPARATE	13.91%	5.16%	2.69%	1.90%	1.13%	0.68%	0.59%	0.34%	0.22%	0.08%
ko+SEPARATE_FC	6.37%	6.75%	4.74%	2.43%	2.00%	1.61%	1.23%	0.22%	0.13%	0.09%

Fig. 68 T438 overall average P&R for baseline and ODC enhanced model

The primary AAPV measures show 3.91% for KO+entire, 3.96% for KO+base, 4.06% for KO+sub, 6.57% for KO+super, 7.25% for KO+separate and 5.95% for KO+separate_FC – demonstrating that ODC search modes achieved better results; although precision was relatively low compared with the other four TREC topics. The secondary outcome was that ontology-based separate document classification provided signification improvement (a factor of 3 times) at the 10% recall point.

6.5.3 Comparing between non-weighted with weighted modes

Search effectiveness comparisons in this group show the effect of using term weighting with the pure language model. The entropy and tf-idf approaches were applied to adjust query weighting. The graph in Fig 69 illustrates overall P&R comparisons for all 10 query sets.



	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
ko+ENTIRE	4.58%	4.10%	3.06%	2.11%	2.01%	1.79%	1.40%	0.61%	0.30%	0.13%
ko_entropy+ENTIRE	4.43%	4.21%	3.25%	2.38%	1.98%	1.82%	1.38%	0.65%	0.30%	0.10%
ko_idf+ENTIRE	4.56%	4.13%	3.20%	2.38%	2.01%	1.87%	1.43%	0.66%	0.30%	0.10%

Fig. 69 T438 overall average P&R for baseline and weighting adjusted model

The results show a negligible improvement was achieved by assigning term weighting, i.e. 3.91% for baseline and 3.96% for both entropy and tf-idf approaches.

6.5.4 Comparison of precision results across all search modes

This subsection assesses search effectiveness outcomes for the 54 search modes. Table 32 represents the search modes without assigned term weighing. Table 33 and Table 34 illustrate the results with entropy weighing and tf-idf weighting respectively. The average precision at the first 30% and 100% recall points, for each search mode, are shown in the last two rows of each table. The primary outcome was that all enhanced search modes provide higher AAPV than the baseline. The results also indicate that

OQE can produce better recall – as presented in individual query search results in Appendix C.

Table 35 shows the merged average first 30% precision results for all 10 query sets. The primary outcome is that SUPER_e+Super provide the best search performance, i.e. 5% AAPV improvement.

To visually compare the performance of each enhanced search mode, Fig. 70 shows relative improvement of search effectiveness over baseline. It indicates that all 53 enhanced search modes produced better precision compared to baseline. The highest improvement was provided by SUPER_e+Super, i.e. 119.01%.

Table 32 T438 Summary of average precision across 10 query sets of Non-weighted modes

T438	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	4.58%	4.31%	4.19%	6.81%	13.91%	6.37%	4.05%	4.85%	4.08%	6.91%	11.42%	7.05%	10.55%	11.28%	11.11%	10.88%	14.98%	10.24%
20%	4.10%	4.19%	3.84%	7.07%	5.16%	6.75%	3.93%	4.33%	3.83%	7.49%	6.55%	5.93%	6.09%	6.42%	7.01%	6.79%	5.71%	5.90%
30%	3.06%	3.38%	4.16%	5.83%	2.69%	4.74%	3.94%	3.75%	4.12%	5.25%	2.92%	5.06%	5.52%	4.22%	5.24%	5.59%	3.10%	4.97%
40%	2.11%	2.53%	3.08%	3.97%	1.90%	2.43%	2.82%	2.79%	3.06%	3.45%	2.01%	2.61%	3.39%	3.27%	3.59%	3.69%	2.24%	2.96%
50%	2.01%	1.88%	2.24%	3.06%	1.13%	2.00%	2.54%	2.06%	2.21%	2.54%	1.21%	2.09%	3.02%	2.75%	2.91%	2.84%	1.48%	2.51%
60%	1.79%	1.61%	1.88%	2.68%	0.68%	1.61%	2.29%	1.85%	1.90%	2.17%	0.67%	1.63%	2.77%	2.28%	2.25%	2.44%	0.94%	1.96%
70%	1.40%	1.12%	1.22%	1.49%	0.59%	1.23%	1.68%	1.46%	1.32%	1.51%	0.52%	1.21%	1.97%	1.63%	1.50%	1.65%	0.64%	1.32%
80%	0.61%	0.57%	0.54%	0.52%	0.34%	0.22%	0.64%	0.56%	0.55%	0.54%	0.37%	0.27%	0.57%	0.48%	0.50%	0.54%	0.37%	0.26%
90%	0.30%	0.18%	0.21%	0.25%	0.22%	0.13%	0.28%	0.19%	0.21%	0.25%	0.22%	0.12%	0.29%	0.21%	0.21%	0.25%	0.22%	0.12%
100%	0.13%	0.09%	0.08%	0.12%	0.08%	0.09%	0.14%	0.09%	0.08%	0.11%	0.08%	0.08%	0.10%	0.12%	0.10%	0.11%	0.08%	0.07%
Avg30	3.91%	3.96%	4.06%	6.57%	7.25%	5.95%	3.97%	4.31%	4.01%	6.55%	6.96%	6.01%	7.39%	7.31%	7.79%	7.75%	7.93%	7.04%
AVG100	2.01%	1.99%	2.14%	3.18%	2.67%	2.56%	2.23%	2.19%	2.14%	3.02%	2.60%	2.61%	3.43%	3.27%	3.44%	3.48%	2.98%	3.03%

Table 33 T438 Summary of average precision across 10 query sets of Entropy-weighted modes

T438	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	4.43%	4.54%	4.61%	7.01%	15.52%	6.65%	4.27%	5.14%	4.58%	7.14%	13.59%	6.75%	10.66%	11.35%	11.27%	11.10%	15.50%	10.34%
20%	4.21%	4.02%	4.08%	7.83%	5.47%	6.27%	4.52%	4.83%	4.04%	8.13%	5.53%	6.18%	7.56%	7.04%	7.62%	8.29%	5.92%	5.47%
30%	3.25%	3.37%	3.79%	6.35%	2.53%	5.31%	3.41%	3.03%	3.73%	5.67%	2.59%	5.31%	5.72%	4.20%	6.17%	6.30%	3.00%	5.46%
40%	2.38%	2.32%	2.93%	3.86%	1.39%	2.71%	2.68%	1.94%	2.89%	3.49%	1.42%	2.77%	3.54%	2.59%	3.70%	3.53%	1.66%	3.18%
50%	1.98%	1.52%	2.22%	2.94%	0.80%	2.00%	2.27%	1.69%	2.20%	2.72%	0.83%	2.03%	3.00%	2.23%	2.93%	2.91%	1.06%	2.41%
60%	1.82%	1.01%	1.81%	2.33%	0.59%	1.37%	2.26%	1.16%	1.81%	2.07%	0.57%	1.43%	2.90%	1.74%	2.48%	2.29%	0.73%	1.68%
70%	1.38%	0.82%	1.12%	1.54%	0.51%	1.07%	1.72%	0.84%	1.22%	1.58%	0.55%	1.06%	2.09%	1.15%	1.67%	1.66%	0.47%	1.10%
80%	0.65%	0.38%	0.45%	0.60%	0.27%	0.27%	0.63%	0.31%	0.45%	0.62%	0.41%	0.35%	0.69%	0.47%	0.61%	0.62%	0.30%	0.30%
90%	0.30%	0.18%	0.21%	0.35%	0.20%	0.12%	0.23%	0.18%	0.21%	0.36%	0.20%	0.12%	0.35%	0.25%	0.28%	0.36%	0.21%	0.12%
100%	0.10%	0.09%	0.08%	0.06%	0.04%	0.10%	0.08%	0.08%	0.08%	0.05%	0.04%	0.09%	0.14%	0.08%	0.08%	0.05%	0.03%	0.09%
Avg30	3.96%	3.97%	4.16%	7.06%	7.84%	6.07%	4.07%	4.33%	4.11%	6.98%	7.24%	6.08%	7.98%	7.53%	8.36%	8.56%	8.14%	7.09%
AVG100	2.05%	1.82%	2.13%	3.29%	2.73%	2.59%	2.21%	1.92%	2.12%	3.18%	2.57%	2.61%	3.67%	3.11%	3.68%	3.71%	2.89%	3.02%

Table 34 T438 Summary of average precision across 10 query sets of tfidf-weighted modes

T438	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
10%	4.56%	4.57%	4.55%	6.82%	15.00%	6.80%	4.22%	5.10%	4.52%	6.91%	13.05%	6.80%	10.70%	11.17%	11.19%	10.92%	14.92%	10.34%
20%	4.13%	4.06%	4.05%	7.91%	5.25%	6.10%	4.18%	4.88%	4.00%	8.11%	5.36%	6.10%	7.30%	7.19%	7.69%	7.51%	5.72%	5.48%
30%	3.20%	3.31%	3.84%	6.29%	2.71%	5.38%	3.49%	3.05%	3.77%	5.79%	2.78%	5.38%	5.79%	4.15%	5.99%	6.33%	3.20%	5.50%
40%	2.38%	2.32%	2.95%	3.78%	1.64%	2.81%	2.67%	1.96%	2.92%	3.50%	1.67%	2.84%	3.54%	2.55%	3.70%	3.54%	1.92%	3.28%
50%	2.01%	1.56%	2.21%	3.00%	0.85%	2.05%	2.29%	1.72%	2.20%	2.65%	0.89%	2.08%	3.07%	2.19%	2.93%	2.84%	1.13%	2.47%
60%	1.87%	1.07%	1.91%	2.43%	0.60%	1.46%	2.28%	1.21%	1.90%	2.20%	0.59%	1.52%	2.98%	1.75%	2.51%	2.43%	0.76%	1.79%
70%	1.43%	0.87%	1.18%	1.54%	0.52%	1.08%	1.73%	0.87%	1.28%	1.57%	0.57%	1.07%	2.17%	1.17%	1.75%	1.66%	0.49%	1.13%
80%	0.66%	0.39%	0.44%	0.62%	0.28%	0.26%	0.64%	0.33%	0.46%	0.63%	0.42%	0.34%	0.67%	0.46%	0.61%	0.64%	0.31%	0.29%
90%	0.30%	0.19%	0.21%	0.35%	0.19%	0.12%	0.23%	0.18%	0.21%	0.35%	0.20%	0.12%	0.35%	0.25%	0.27%	0.35%	0.20%	0.12%
100%	0.10%	0.09%	0.08%	0.06%	0.05%	0.10%	0.08%	0.08%	0.08%	0.05%	0.05%	0.09%	0.16%	0.08%	0.08%	0.05%	0.05%	0.09%
Avg30	3.96%	3.98%	4.14%	7.01%	7.65%	6.09%	3.96%	4.35%	4.10%	6.93%	7.07%	6.09%	7.93%	7.50%	8.29%	8.25%	7.95%	7.11%
AVG100	2.06%	1.84%	2.14%	3.28%	2.71%	2.62%	2.18%	1.94%	2.13%	3.18%	2.56%	2.63%	3.67%	3.10%	3.67%	3.63%	2.87%	3.05%

Table 35 T438 Average 30% APV of 54 search modes

T438	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Baseline	3.91%	3.96%	4.06%	6.57%	7.25%	5.95%	3.97%	4.31%	4.01%	6.55%	6.96%	6.01%	7.39%	7.31%	7.79%	7.75%	7.93%	7.04%
Entropy	3.96%	3.97%	4.16%	7.06%	7.84%	6.07%	4.07%	4.33%	4.11%	6.98%	7.24%	6.08%	7.98%	7.53%	8.36%	8.56%	8.14%	7.09%
Tf-Idf	3.96%	3.98%	4.14%	7.01%	7.65%	6.09%	3.96%	4.35%	4.10%	6.93%	7.07%	6.09%	7.93%	7.50%	8.29%	8.25%	7.95%	7.11%

Table 36 T438 improvments

T438	KO+ entire	KO+ base	KO+ sub	KO+ super	KO+ sep	KO+ sep_fc	SUB+ entire	SUB+ base	SUB+ sub	SUB+ super	SUB+ sep	SUB+ sep_fc	SUPER +entire	SUPER +base	SUPER +sub	SUPER +super	SUPER +sep	SUPER+ sep_fc
Non-w	0.00%	1.34%	3.94%	68.05%	85.53%	52.22%	1.62%	10.17%	2.63%	67.60%	78.07%	53.80%	88.93%	86.83%	99.13%	98.27%	102.88%	80.00%
entropy	1.26%	1.64%	6.41%	80.52%	100.45%	55.35%	4.01%	10.85%	5.22%	78.53%	85.10%	55.50%	104.04%	92.61%	113.71%	119.01%	108.17%	81.30%
Tf-Idf	1.38%	1.83%	5.95%	79.16%	95.72%	55.77%	1.35%	11.14%	4.73%	77.35%	80.72%	55.82%	102.79%	91.88%	112.01%	111.03%	103.29%	81.76%
AVG	0.88%	1.60%	5.43%	75.91%	93.90%	54.44%	2.33%	10.72%	4.19%	74.49%	81.30%	55.04%	98.59%	90.44%	108.28%	109.44%	104.78%	81.02%

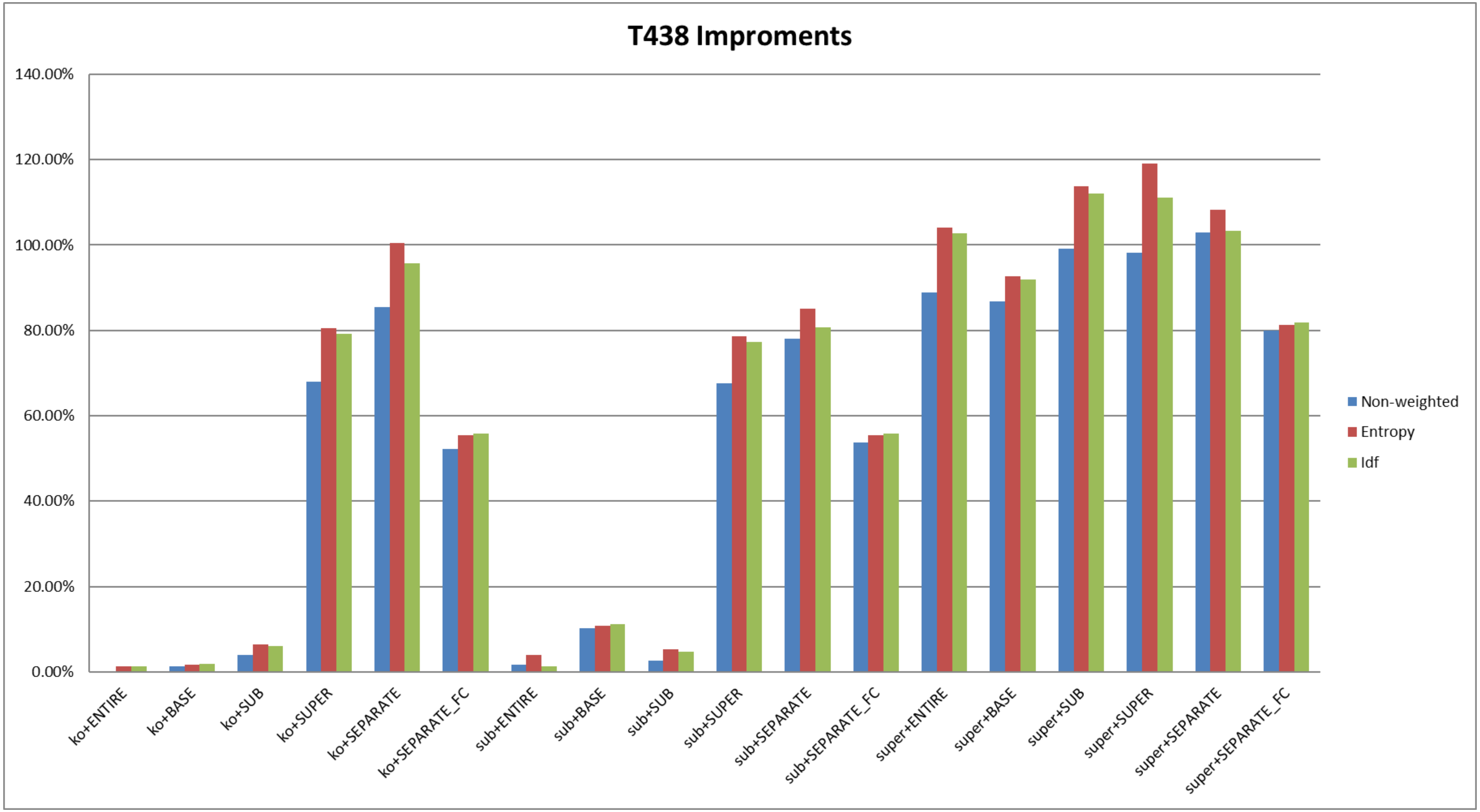


Fig. 70 T438 search mode improvements

6.5.5 Reflection on Hypotheses

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

Whilst the tourism ontology was not designed specifically for T438, it is tourism related. T438's extended and wider, multi-context ontology achieved inconclusive results, probably because a more general ontology would be more likely to generate ambiguous terms. T438 experiment results in subsection 6.5.1 show that OQE enhanced search modes produced higher AAPV against keyword-only search. The AAPV of SUPER+entire was approximately 88.93% higher than baseline; whilst, the analysis of individual query results in appendix C, shows overall recalls were better than keyword only modes

The above outcomes demonstrate the benefit of applying the query expansion beyond baseline and provide clear evidence to support the hypothesis, for both precision and recall.

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

As in the previous two topic experiments, the entropy and tf-idf algorithms were used for term weighting, to eliminate possible, ambiguous expanded terms. The results in subsection 6.5.3 show that assigning lower weightings to expansion terms can, to some extent, enhance retrieval performance. It is considered that the hypothesis has been supported.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or classification based local smoothing strategies are more effective than the global strategies. "*

The ontology based document classification (Base, Sub, Super, Separate and Separate_FC) provides the corpus structure for local smoothing. The benefit of extended document classification was shown by the experiment results, i.e. KO+separate achieved about 85.53% improvement against the baseline. As Fig 68 showed in subsection 6.5.2, all ODC enhanced search modes demonstrated signification improvement over keyword only search mode. Overall, this provides good evidence to support the hypothesis.

7. EVALUATION

This chapter evaluates the 5 search topic experiments, by first proposing a summary of experiment results and then, by way of a critical review of the work undertaken.

7.1 Summary of experiment results

This subsection evaluates primary experiment outcomes from two perspectives, i.e. AAPV measure and P&R outcomes.

7.1.1 Performance outcomes using AAPV measures

Search effectiveness was primarily based on average precision percentage values of first 30% recall point (the AAPV). The AAPVs of 54 search modes used in 5 topic experiments are shown in Fig 71, with Fig 72 providing the average improvement percentages against the baseline.

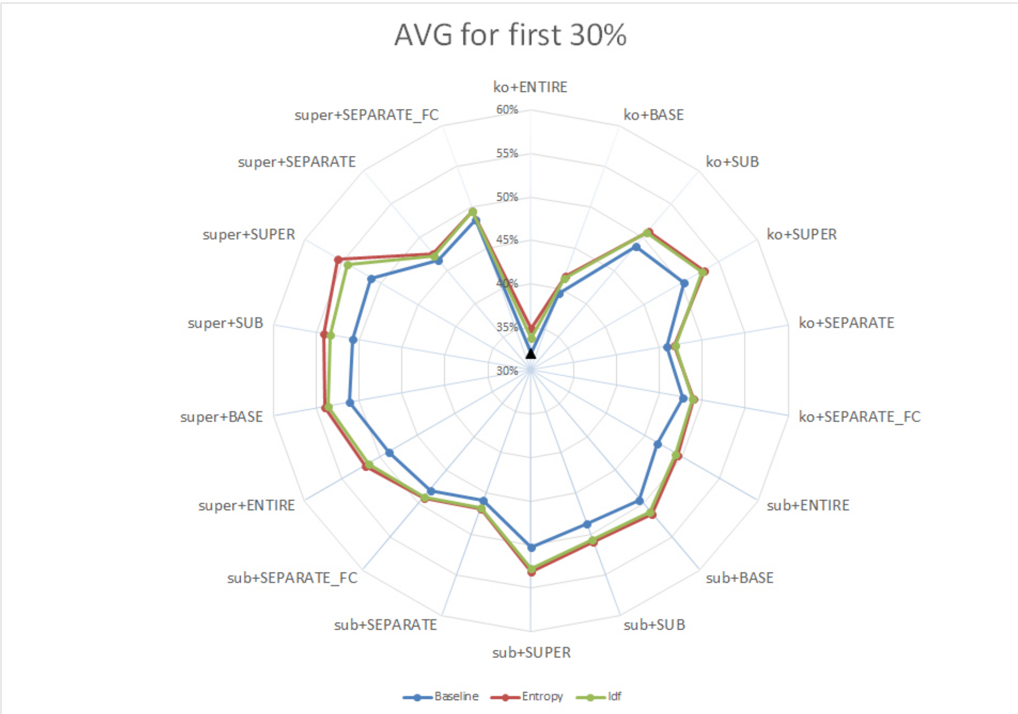


Fig. 71 Five experiments average percentage search effectiveness first 30% recall

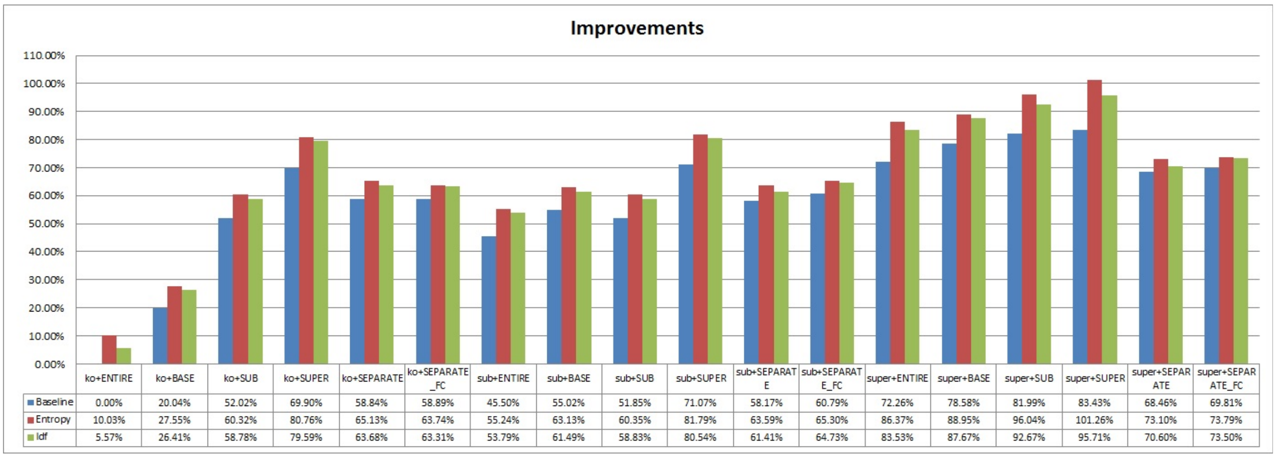


Fig. 72 Bar chart comparison of search effectiveness over baseline

In Fig 71, the baseline is denoted by the small black triangle. Measurements show that the established 53 search modes achieved remarkable search effectiveness improvements. The Bar chart in Fig 72 demonstrates consistently similar results, especially in the four SUPER_e+super modes, which achieved the best improvement in all 54 search experiments, i.e. 101.26% improvement.

The following three radar graphs show the average precisions in 10%, 20% and 30% recall points for all 50 experiment queries, i.e. 10 query sets for each of 5 search experiments.

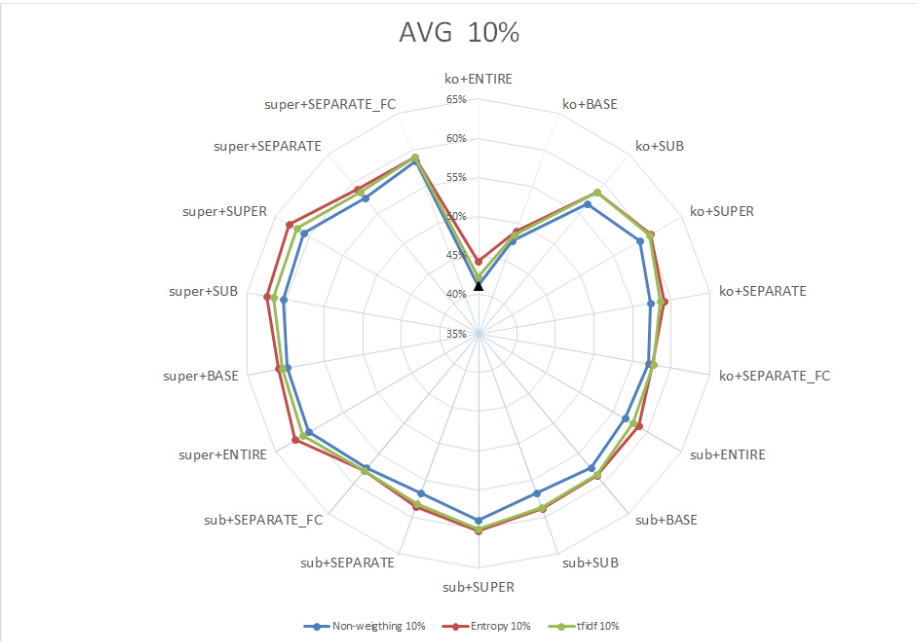


Fig. 73 Five experiment average precisions at 10% recall point

The primary outcomes were that the 53 search modes all achieved better AAPVs when compared to the baseline. The results also indicated that the average precisions of sub class with entropy weight-enhanced search models (red line) were better than non-weighted (blue line) and tf-idf weighted enhanced models (green line).

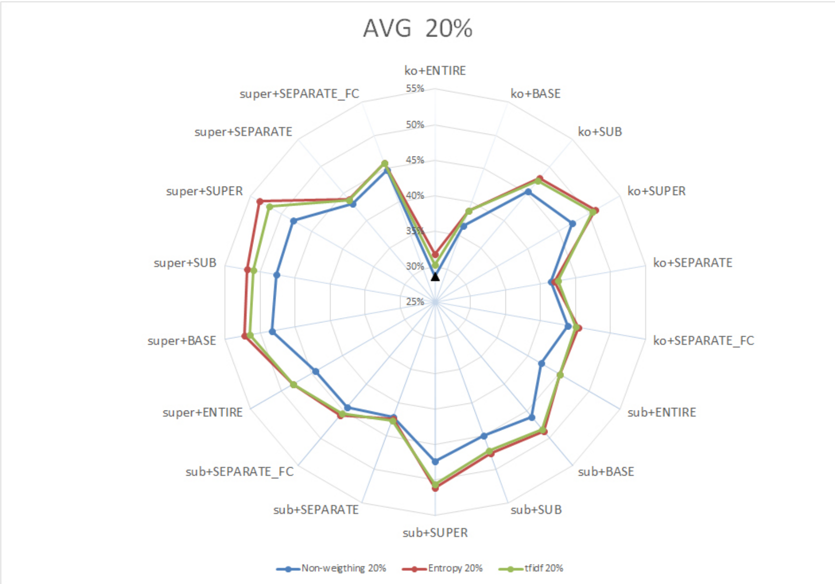


Fig. 74 Five experiment average precisions at 20% recall point

Fig 74 demonstrates that the ontology enhanced search model provides a good AAPV improvement in search effectiveness over baseline. The term weight-assigned mode is remarkable better than the non-weighted mode. The secondary outcome is that SUPER_e+super is the best search mode in this recall interval.

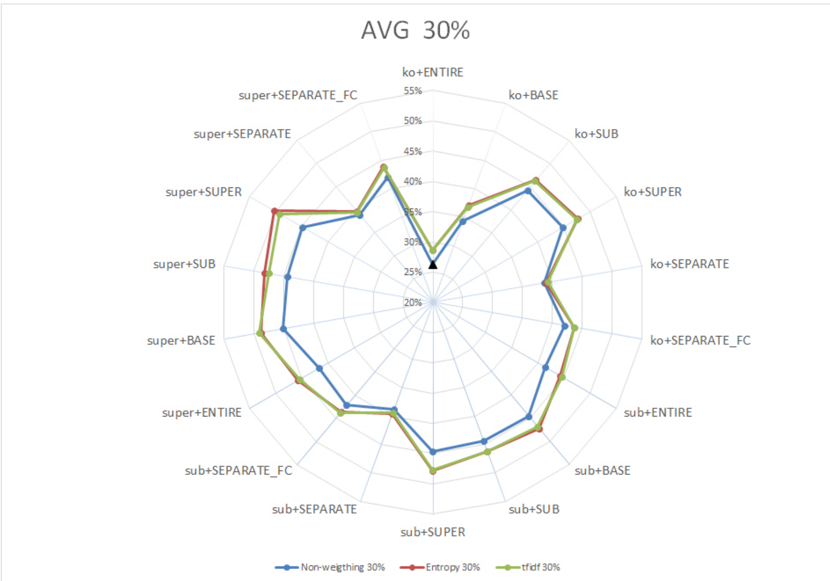


Fig. 75 Five experiment average precisions at 30% recall point

The results of average precisions at 30% recall point, in Fig 75, show similar results to the previous two recall interval graphs. All established search modes achieved marked search effectiveness improvement. Entropy weighting still produced better results than non-weighted modes and tf-idf enhanced modes.

7.1.2 Precision success and recall outcomes

Table 37 summarizes the percentage of times that baseline and ontology enhanced model (OEM) produced the highest precision success at the first 30% recall points and shows the overall recall performance in the topic experiments. The results reflect the similar outcomes to the AAPV performance in 6.1.1 and confirm the search effectiveness improvements in both precision and recall.

Table 37 Comparison of 5 experiments Ontology enhanced search model success

TREC Topic	Precision Success			Recall Success		
	% Baseline top	% OEMs top	%Tied	% Baseline top	% OEMs top	%Tied
T401	0%	90%	10%	0%	0%	100%
T403	0%	100%	0%	0%	60%	40%
T416	0%	90%	10%	0%	40%	60%
T431	0%	100%	0%	0%	40%	60%
T438	0%	100%	0%	0%	10%	90%

The summary shows that OEMs provided better search effectiveness compared to the baseline. In particular, the two non-specific topics achieved marked improvements, i.e. all T431 Robotic Technology and T438 Tourism achieved improvement in 100% of queries. Individual topic outcomes were:

- T401 OEM AAPV showed highly improved the search effectiveness compared to baseline, i.e. 90% of queries achieved positive results, with 10% of T401 queries showing tied outcomes. There was no recall improvement in 10 queries.

- T403 OEMs delivered better search performance; all 10 queries had better AAPVs against the baseline, and only one of them produced a poorer result. OEMs also produced improved recall, i.e. 60% queries achieved better recall.

- T416 demonstrated that OEMs can provide improved precision, with 90% of query searches improved and 10% of them achieved tied outcomes; they also showed better recall compared to baseline, i.e. with 40% of queries improved and 60% tied.

- T431 OEMs produced strong search performance in both precision and recall, i.e. 100% search queries had improved results, and 10% queries achieved better recall.

- T438 OEMs were clearly the most effective search experiment, across 100% of queries and with 10% of queries achieving better recalls.

The measurements demonstrate that, in most of the search experiments, OEM achieved better results compared with baseline, with 48 of 50 queries achieving positive results and 2 of them achieving tied results.

7.2 Critical Review

i. The outcomes from AAPV (6.1.1) and P&R (6.1.2) subsection have clearly demonstrated the benefit of using an ontology enhanced search model, using small ontology contexts, to improve search effectiveness.

APVs serve as a consistent and primary measure of search performance, with the focus on early recall points seeking to simulate that a typical Web user might be mainly interested high precision (first few relevant pages) of search engine results. Based on P&R outcomes, the APV measures have verified the benefit of OEMs search mode. Overall recalls (using average precision in all recall points) were demonstrated in Chapter 5 and showed markedly improved precision across the entire recall range. The outcomes of T403, T416, T431 and T438 also showed that the OQE enhanced search mode, with sub class and super class expansion, can achieve a higher recall compared with baseline.

Precision and recall successes showed that non-specific topic experiments (T431 Robotic-technology and T438 Tourism) demonstrated greater precision improvements compared to the specific topics (T401 Immigration, T403 Osteoporosis and T416 Three Gorges Project). The public set of ontology contexts provided wider, accurate and potentially relevant query terms for searching.

ii. The experiment processes required several manually developed modular ontologies for use with the predetermined TREC topics. A fully resourced search engine would need a vast set of ontologies to cover different and specific domains. This presents a challenge, to develop a vast resource of ontologies, which would be a significant “cost” compared to the benefit of exploiting ontology in an IR system.

iii. A small and specialised ontology offers more specific domain vocabularies, to narrow the search area. The average precisions of first three topics with specialised ontologies (83.05%, 55.93% and 57.93%) were much higher than T431 and T438’s (12.08% and 7.41%) non-specialised ontologies.

iv. The experiments use entropy and tf-idf weighting to control the quality of terms. All five topic experiments demonstrated remarkable improvements compared with non-weighted modes, with entropy achieving better performance than tf-idf weighting.

v. In the five search experiments, all of the established modes produced better AAPV in all 50 queries; these results appear to suggest that SUPER_e+super is the best search mode of all. Only in T401 search experiments, the best approach was SUPER_e+separate, which may have been caused by the relatively shallow ontology structure. Immigration ontology has only 32 classes, where 11 of them are single classes (i.e. no subclass or superclass), with the result that Immigration’s OQE and ODC expansion capacity was much lower than the other four ontologies. T401 is the only topic that used a locally built ontology because no suitable public ontology is available.

vi. The following comments refer to the research hypotheses proposed in subsection 2.5. For consistent evaluation of effectiveness by applying OQE and ODC to the language model, hypothesis (i) comments focus on the OQE search process with the entire collection smoothing, hypothesis (ii) concentrates on the impact of applying term weighting to SLM and OEMs, and hypothesis (iii) focuses on those modes only enhanced by ODC (KO+base, KO+sub, KO+separate).

Hypothesis (i) - *"Topic specific ontology context based OQE can have a positive impact on precision and recall. Query term-matched classes may have more beneficial "terminology" beyond simply expanding the keyword by using "general terms". Exploiting ontology will provide useful OQE options and improve document relevance scoring and ranking."*

T401, T403 and T416 exploited topic-specific ontologies for the OQE process. The experiment results demonstrated in Chapter 5 highlighted markedly improved precision when using specific ontologies: T401 Immigration SUPER+entire model provided 13.23% improvement over baseline in 9 search queries; similarly, T403's more specific bone health ontology produced the most search effectiveness improvements, i.e. 71.23% improvement. T416, SUPER+entire produced 133.00% improvement against baseline. The AAPV outcomes support the first part of the hypothesis *"Topic specific small ontology context based OQE can have a positive impact on precision and recall."*

T431 Robotic ontology contained robotic related terms and relationships for OQE and ODC process, but was not specifically designed for T431 robotic application searching. Also, the tourism ontology was not designed specifically for T438, although it was associated with tourism. As a result, T438 provided a wider and multi-context ontology, which can easily generate general or ambiguous terms. The T431 experiment results in subsection 6.4.1 show that the OQE enhanced search model produced higher AAPV against keyword-only search. The AAPV of SUPER+entire was approximately 58.43%

higher than baseline and, based on the individual query results analysis in appendix C, overall recalls were better than keyword only modes

These outcomes demonstrate the benefit of applying the query expansion beyond baseline, and provide clear evidence to support the hypothesis, in terms of both precision and recall.

Hypothesis (ii) - *"Assigning lower weighting to expanded terms can enhance the retrieval accuracy in the OQE process. Term weighting reflects the distance between the initial query and expanded terms. "*

The experiment used two different weighting approaches to minimize the problem of query expansion term ambiguity. The overall results show that assigning term weighting can produce better results compared with non-weighted modes. The average APV for 18 non-weighted modes is 47.21%, 49.44% for 18 entropy weighting assigned modes and 49.08% for 18 tf-idf adjusted search modes.

The results clearly reflect that assigning weighting to query terms can enhance the retrieval accuracy. Term weighting could help a system to eliminate an “over-biased” query expansion problem. Overall, the experiment results support the hypothesis.

Hypothesis (iii) - *"Exploiting the corpus structure in the language model smoothing can provide more accurate smoothing of language model. Document clustering or classification based local smoothing strategies are more effective than the global strategies. "*

Ontology produces a semantic-based document classification approach to apply in local smoothing. The benefits of document classification were indicated in the 5 experiment results, with 13.23% AAPV improvement for

T401 (KO+separate), 76.87% for T403 (KO+super), 85.53% for T416 (KO+separate), 35.64% for T431 (KO+separate) and 165.06% for T438 (KO+super). All ODC enhanced search models demonstrated search performance improvement to differing extents. However, the document classification approach that is the best for SLM remains uncertain. Therefore, additional experimentation would be beneficial, to further test the best ODC approach.

Overall, ontology based document classification has produced a more accurate corpus structure for local smoothing and achieved positive results. It is considered that the hypothesis has been proved.

8 CONCLUSIONS & FUTURE WORK

The objective of this PhD was to provide a novel approach to improving retrieval performance, by applying Ontology to a statistical language model (SLM). The proposed methods consisted of two major processes, namely ontology-based query expansion (OQE) and ontology-based document classification (ODC) enhanced smoothing (eS). This chapter draws conclusions from the project outcomes and discusses future directions of exploration.

8.1 CONCLUSION

In this research, we proposed an ontology enhanced SLM for ad hoc text retrieval. The essential task of SLM-based information retrieval is to accurately estimate the document model, which can also include how to solve the data sparseness problems, i.e. when a search term does not occur in a document, the probability of the document related to the topic will be zero. This unreadable zero probability typically was assigned by maximum likelihood estimation, based on the context of a document, which is typically affected by insufficient sampling; a smoothing process was proposed to address this problem.

A traditional language model estimates the probability of an unseen word, based on its probability distribution in the entire collection, to smooth the document model. In this project, we provided a novel approach to improve search effectiveness (precision and recall) by applying ontology context to enhance language model smoothing. The proposed 53 enhanced search modes consisted of two major processes, OQE and ODC. The proposed search model not only gave a bigger sampling, to estimate the document model, but also provided accurate local probability distribution using a document categories feature.

The search experiment outcomes have justified the approached adopted. Empirical results have shown that search effectiveness of 53 enhanced search modes improved over the basic language model. The subsection 7.1.2 average P&R results show that in a total of 50 query term sets of 5 experiments, all ontology enhanced language models provided improved APV. The results also highlight that 30% of query sets achieved improved recall.

In section 7.2, the hypotheses were fully considered, based on the search experiment outcomes. Experiment results provided good evidence to support: hypothesis (i), i.e. specific ontology context based query expansion can improve search precision and recall; hypothesis (ii) either assigned entropy based term weighting or tf-idf based term weighting can provide small search effectiveness improvements compared with baseline. The search outcomes have also provided good evidence to support hypothesis (iii) that the AVP of ontology-based document classification enhanced SLMs are all better than the pure language model with entire collection smoothing. Moreover, the ODC experiments provided a useful solution, when combined with an OQE process, to improve the precision.

Overall, the LMST engine has successfully achieved the primary objective of improving search effectiveness. Topic specific ontology contexts are worthwhile for exploiting the SLM, and the best APV results can be achieved by using combinations of OQE and ODC processes.

8.2 FUTURE WORK

The ontology enhanced search model we proposed is just an initial step to understanding query expansion and document clustering/classification based smoothing. There are many potential future research directions, following the discussion in this thesis, as follows:

i. **Beyond bag of words**

   For this project, we would follow the existing research efforts that currently simplify text documents with a bag of words, where the order of words in document is ignored. One potential research direction could be to represent the document context using more flexible level of language units, e.g. bigram or trigram. A focus on the language unit could provide even more research opportunity, e.g. how to provide the ontology based query expansion to the language unit, or, how to calculate the term weighting for the proximate words?

ii. **Ontology refinement**

As previously discussed in section 5.3, T401, T416 experiments used simple ontology contexts with flat hierarchy (less than 4 layers), which provided only limited potential for query term expansion. One important direction would be to explore the existing ontology in specific domains with more complex and accurate structures. Professional ontologies can provide rich prior knowledge and understanding over the content in these domains; thereby using these ontologies could further improve search performance. To apply such extensive ontologies, with wider ranging and deeper hierarchies, the ontology traversal algorithm will also require further research, to determine the optimum number of concept levels to include in an OQE.

iii. Different document clustering/classification

The experiment outcomes demonstrate the remarkable search effectiveness improvement using the ODC-only enhanced search model, i.e. KO+base, KO+sub, KO+super, KO+separate and KO+separate_FC. While OQE can make a positive impact towards search effectiveness improvements, the search results appear to suggest that, as a supplemental support, combining OQE with robust document corpus structure can achieve even better outcomes. Therefore, different document classification/clustering approaches are worth exploiting, combined with OQE process, such as Latent Semantic Indexing, graph structure, and document expansion.

iv. More Ontology and database testing

In this research, because generating modular ontologies is time consuming, only 5 topics were selected for the search performance evaluation experiments. Further, different types of existing ontologies would be worth exploiting in OQE and ODC testing. At the same time, using other specialized databases would be a potential good direction for future work, such as medicine database or archaeology database.

v. Smoothing parameter training

As previously mentioned in subsection 5.3, the standard approach for determining the Dirichlet smoothing parameter is from the training data which consists of independent queries and relevance judgments. It is hard to find training data that has appropriate topic and relevance judgments, so comparison experiments selected a μ value that can maximize the search effectiveness. Instead of comparing the theoretical optimum search effectiveness for both the baseline and enhanced model, one potential research direction could be to use an unsupervised approach to estimate the smoothing parameters.

vi. Different term weighting

In the research, the experiment results suggest that assigning different term weighing will directly affect the search performance. Therefore one potential direction could be to experiment with more term weighting approaches in the query expansion process, to improve the SLM based IR, e.g. BM25, semantic-based context weighting.

References

AKAIKE, H. (1973). "Information theory and an extension of the maximum likelihood principle." Budapest, Hungary, pp. 267–281.

ANDREOU, A. (2005). "Ontologies and Query expansion." Master of Science, School of Informatics, University of Edinburgh.

ANTONIOU, G.&HARMELEN, F. (2009). "Web ontology language: Owl." Handbook on ontologies, pp. 91-110.

BAEZA-YATES, R.&RIBEIRO-NETO, B. (1999). Modern information retrieval, ACM press New York.

BAO, J., CAO, Y., et al. (2004). Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. IKE.

BECHHOFER, S., VAN HARMELEN, F., et al. (2004). "OWL web ontology language reference." W3C recommendation 10.

BERGER, A.&LAFFERTY, J. (1999). Information retrieval as statistical translation. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. Berkeley, California, United States, ACM, pp. 222-229.

BERNERS-LEE, T. (1998). "Semantic web road map."

BERNERS-LEE, T. (2006). "Linked data." International Journal on Semantic Web and Information Systems 4(2).

BERNERS-LEE, T., FIELDING, R., et al. (1998). "Uniform resource identifiers (URI): generic syntax."

BERNERS-LEE, T., LASSILA, O., et al. (2001). "The semantic web." Scientific American 284(5), pp. 34-43.

BILLERBECK, B.&ZOBEL, J. (2004). Questioning query expansion: An examination of behaviour and parameters. Proceedings of the 15th Australasian database conference-Volume 27, Australian Computer Society, Inc.

BLANCO, R.&BARREIRO, A. (2008). Probabilistic document length priors for language models. Advances in Information Retrieval, Springer, pp. 394-405.

BORGO, S., GANGEMI, A., et al. (2002). "WonderWeb Deliverable D15 Ontology RoadMap." The WonderWeb Library of Foundational Ontologies and the DOLCE ontology, pp. 08-15.

BRAUNSCHWEIG, B.&RAINAUD, J. (2005). "Semantic web applications for the petroleum industry [Online]." Avaliable from http://www.eswc2005.org/documents/ESWC2005_Braunschweig1_41.pdf. [Accessed 09 December 2010].

BRICKLEY, D.&MILLER, L. (2005). "FOAF vocabulary specification [Online]." Avaliable from http://xmlns.com/foaf/spec/20071002.html. [Accessed 10 December 2010].

BUCKLEY, C., SALTON, G., et al. (1995). "Automatic query expansion using SMART: TREC 3." NIST special publication sp, pp. 69-69.

CHANDRASEKARAN, B., JOSEPHSON, J., et al. (2002). "What are ontologies, and why do we need them?" Intelligent Systems and Their Applications, IEEE 14(1), pp. 20-26.

CHEN, H., LYNCH, K. J., et al. (1993). "Generating, integrating, and activating thesauri for concept-based document retrieval." IEEE Expert 8(2), pp. 25-34.

CHEN, S. F.&GOODMAN, J. (1996). An empirical study of smoothing techniques for language modeling. Proceedings of the 34th annual meeting on Association for Computational Linguistics, Association for Computational Linguistics.

CHU, W. W., LIU, Z., et al. (2002). "Textual document indexing and retrieval via knowledge sources and data mining." Communication of the Institute of Information and Computing Machinery (CIICM), Taiwan 5(2).

CORSARO, W. A. (1982). "Something Old and Something New The Importance of Prior Ethnography in the Collection and Analysis of Audiovisual Data." Sociological Methods & Research 11(2), pp. 145-166.

CROFT, W. (1993). "Knowledge-based and statistical approaches to text retrieval." IEEE Expert 8(2), pp. 8-12.

CROFT, W. B. (2003). Language models for information retrieval. Data Engineering, 2003. Proceedings. 19th International Conference on, IEEE.

CUI, H., WEN, J.-R., et al. (2003). "Query expansion by mining user logs." Knowledge and Data Engineering, IEEE Transactions on 15(4), pp. 829-839.

D AZ-GALIANO, M. C., MART N-VALDIVIA, M. T., et al. (2009). "Query expansion with a medical ontology to improve a multimodal information retrieval system." Computers in Biology and Medicine 39(4), pp. 396-403.

DECKER, S., MITRA, P., et al. (2000). "Framework for the semantic Web: an RDF tutorial." Internet Computing, IEEE 4(6), pp. 68-73.

DING, L., FININ, T., et al. (2004). "Swoogle: a search and metadata engine for the semantic web." ACM Trans. Inf. Syst., pp. 652-659.

FINKELSTEIN, L., GABRILOVICH, E., et al. (2001). Placing search in context: The concept revisited. Proceedings of the 10th international conference on World Wide Web, ACM.

FU, G., JONES, C. B., et al. (2005). Ontology-based spatial query expansion in information retrieval. On the move to meaningful internet systems 2005: CoopIS, DOA, and ODBASE, Springer, pp. 1466-1482.

FURNAS, G., LANDAUER, T., et al. (1987). "The vocabulary problem in human-system communication." Communications of the ACM 30(11), pp. 964-971.

GEORGE, D. (2010). Examining the Application of Modular and Contextualised Ontology in Query Expansions for Information Retrieval, University of Central Lancashire. PHD.

GONDY, L., TOLLE, K., et al. (1999). Customizable and Ontology-Enhanced Medical Information Retrieval Interfaces. Proceedings of IMIA Working Group.

GONZALO, J., VERDEJO, F., et al. (1998). "Indexing with WordNet synsets can improve text retrieval." arXiv preprint cmp-lg/9808002.

GRUBER, T. (1993). "A translation approach to portable ontology specifications." Knowledge acquisition 5, pp. 199-199.

GUARINO, N. (1998). Formal Onthology in Information Systems: Proceedings of the First International Conference (FIOS'98), June 6-8, Trento, Italy, IOS press.

GUIZZARDI, G. (2007). On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta)Models. Proceeding of the 2007 conference on Databases and Information Systems IV: Selected Papers from the Seventh International Baltic Conference DB\&IS'2006, IOS Press, pp. 18-39.

HARMAN, D. (1992). Relevance feedback revisited. Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

HIEMSTRA, D. (2001). Using Language Models for Information Retrieval. Enschede, pp. 164.

HIEMSTRA, D. (2009). "Information retrieval models." Information Retrieval: searching in the 21st Century, pp. 1-19.

HORRIDGE, M.&TSARKOV, D. (2006). "Supporting early adoption of OWL 1.1 with Protege-OWL and FaCT++ [Online]." Avaliable from http://www.webont.org/owled/2006/acceptedLong/submission_15.pdf. [Accessed 13 October 2010].

HORROCKS, I., FENSEL, D., et al. (2002). "The ontology inference layer OIL [Online]." Avaliable from http://www. ontoknowledge. org/oil/papers. shtml. [Accessed 5 April 2011].

JOHO, H., SANDERSON, M., et al. (2004). A study of user interaction with a concept-based interactive query expansion support tool. Advances in Information Retrieval, Springer, pp. 42-56.

JONES, S. (1993). "A thesaurus data model for an intelligent retrieval system." Journal of Information Science 19(3), pp. 167-178.

JONES, S., GATFORD, M., et al. (1995). "Interactive Thesaurus navigation: intelligence rules OK?" Journal of the American Society for Information Science 46(1), pp. 52-59.

KROVETZ, R.&CROFT, W. B. (1992). "Lexical ambiguity and information retrieval." ACM Transactions on Information Systems (TOIS) 10(2), pp. 115-141.

KURLAND, O.&LEE, L. (2004). "Corpus structure, language models, and ad hoc information retrieval [Online]." Avaliable from http://www.cs.cornell.edu/home/llee/papers/clustir.pdf. [Accessed 12 December 2013].

KURLAND, O.&LEE, L. (2005). "PageRank without hyperlinks: Structural re-ranking using links induced by language models [Online]." Avaliable from http://www.cs.cornell.edu/home/llee/papers/lmpagerank.pdf. [Accessed 10 October 2013].

LAFFERTY, J.&ZHAI, C. (2001). "Document language models, query models, and risk minimization for information retrieval [Online]." Avaliable from http://ciir.cs.umass.edu/irchallenges/risk.pdf. [Accessed 17 December 2013].

LANDAUER, T. K., MCNAMARA, D. S., et al. (2013). Handbook of latent semantic analysis, Psychology Press.

LAVRENKO, V., CHOQUETTE, M., et al. (2002). "Cross-lingual relevance models [Online]." Avaliable from http://delivery.acm.org/10.1145/570000/564408/p175-lavrenko.pdf?key1=564408&key2=7314907921&coll=DL&dl=ACM&CFID=9110461&CFTOKEN=57910532. [Accessed 08 March 2013].

LENAT, D. B. (1995). "CYC: A large-scale investment in knowledge infrastructure." Communications of the ACM 38(11), pp. 33-38.

LIU, X.&CROFT, W. B. (2004). Cluster-based retrieval using language models. Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

LV, Y.&ZHAI, C. (2009). Positional language models for information retrieval. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, ACM.

MAEDCHE, A.&STAAB, S. (2005). "Ontology learning for the semantic web." Intelligent Systems, IEEE 16(2), pp. 72-79.

MAGUIRE, E. A., FRITH, C. D., et al. (1999). "The functional neuroanatomy of comprehension and memory: the importance of prior knowledge." Brain 122(10), pp. 1839-1850.

MANNING, C. D., RAGHAVAN, P., et al. (2008). "Introduction to information retrieval."

MANOLA, F., MILLER, E., et al. (2004). "RDF primer." W3C recommendation 10.

MARCU, D. (1997). "The rhetorical parsing of natural language texts [Online]." Avaliable from http://delivery.acm.org/10.1145/980000/979630/p96-marcu.pdf?key1=979630&key2=3814907921&coll=DL&dl=ACM&CFID=9110461&CFTOKEN=57910532. [Accessed 06 March 2013].

MATTHEWS, B. (2005). "Semantic Web Technologies." E-learning 6(6), pp. 8.

MCCALLUM, A.&NIGAM, K. (1998). "A comparison of event models for naive Bayes text classification." Workshop on Learning for Text Categorization.

MEI, Q., ZHANG, D., et al. (2008). A general optimization framework for smoothing language models on graph structures. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

MILLER, G. A. (1995). "WordNet: a lexical database for English." Communications of the ACM 38(11), pp. 39-41.

MITRA, M., SINGHAL, A., et al. (1998). Improving automatic query expansion. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

NICHTICH (2008). "Recall & Precision [Online]." Avaliable from http://en.wikipedia.org/wiki/Precision_and_recall. [Accessed 2 May 2013].

NILSSON, K., HJELM, H., et al. (2011). SUiS–cross-language ontology-driven information retrieval in a restricted domain. Proceedings of the 15th NODALIDA conference, Citeseer.

NOY, N. F.&HAFNER, C. D. (1997). "The state of the art in ontology design: A survey and comparative review." AI magazine 18(3), pp. 53.

PARALIC, I. K. A. J. (2003). "Ontology-based Information Retrieval." Information and Intelligent Systems, Croatia, pp. 23-28.

PONTE, J. M.&CROFT, W. B. (1998). A language modeling approach to information retrieval. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. Melbourne, Australia, ACM, pp. 275-281.

PRABHAKAR RAGHAVAN , H. S. A. C. D. M. (2008). Introduction to Information Retrieval, Cambridge University Press.

PUSTEJOVSKY, J. (1995). "The core lexical engine: The contextual determination of word sense [Online]." Avaliable from http://www.cs.tufts.edu/~jacob/isgw/Pustejovsky.html. [Accessed 10 October 2013].

RIJSBERGEN, C. J. V. (1979). Information Retrieval, Dept. of Computer Science, University of Glasgow.

ROBERTSON, A.&WILLETT, P. (1993). "A Comparison of Spelling-Correction Methods for the Identification of Word Forms in Historical Text Databases." Literary and Linguistic Computing 8(3), pp. 143-152.

ROBERTSON, S. E. (1977). The probability ranking principle in IR. Readings in information retrieval, Morgan Kaufmann Publishers Inc., pp. 281-286.

ROBERTSON, S. E. (1990). "On term selection for query expansion." Journal of Documentation 46(4), pp. 359-364.

ROCCHIO, J. J. (1971). "Relevance feedback in information retrieval."

RUTHVEN, I.&LALMAS, M. (2003). "A survey on the use of relevance feedback for information access systems." The Knowledge Engineering Review 18(02), pp. 95-145.

SALTON, G.&MCGILL, M. J. (1983). "Introduction to modern information retrieval."

SALTON, G., WONG, A., et al. (1975). "A vector space model for automatic indexing." Commun. ACM 18(11), pp. 613-620.

SANDERSON, M.&LAWRIE, D. (2000). Building, testing, and applying concept hierarchies. Advances in information retrieval, Springer, pp. 235-266.

SI, L., JIN, R., et al. (2002). "A language modeling framework for resource selection and results merging [Online]." Avaliable from http://delivery.acm.org/10.1145/590000/584856/p391-si.pdf?key1=584856&key2=2724907921&coll=DL&dl=ACM&CFID=9110461&CFTOKEN=57910532. [Accessed 20 September 2013].

SIHVONEN, A.&VAKKARI, P. (2004). "Subject knowledge improves interactive query expansion assisted by a thesaurus." Journal of Documentation 60(6), pp. 673-690.

SONG, F.&CROFT, W. B. (1999). A general language model for information retrieval. Proceedings of the eighth international conference on Information and knowledge management, ACM.

SP RCK JONES, K. (1997). Readings in in Information Retrieval, Morgan Kaufmann.

SP RCK JONES, K., WALKER, S., et al. (2000). "A probabilistic model of information retrieval: development and comparative experiments: Part 1." Information Processing & Management 36(6), pp. 779-808.

STUDER, R., BENJAMINS, V., et al. (1998). "Knowledge engineering: principles and methods." Data & knowledge engineering 25(1-2), pp. 161-197.

TOMBROS, A.&SANDERSON, M. (1998). Advantages of query biased summaries in information retrieval. Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

TSARKOV, D.&HORROCKS, I. (2006). "FaCT++ description logic reasoner: System description." Automated Reasoning, pp. 292-297.

ULLMANN, S. (1972). Semantics: an introduction to the science of meaning, Blackwell.

UWE M. BORGHOFF, R. P. (1998). Information technology for knowledge management, Springer Verlag.

VAKKARI, P., JONES, S., et al. (2004). "Query exhaustivity, relevance feedback and search success in automatic and interactive query expansion." Journal of Documentation 60(2), pp. 109-127.

VALLET, D., FERNANDEZ, M., et al. (2005). An Ontology-Based Information Retrieval Model. The Semantic Web: Research and Applications.

VECHTOMOVA, O., ROBERTSON, S., et al. (2003). "Query expansion with long-span collocates." Information Retrieval 6(2), pp. 251-273.

VOORHEES, E. M. (1993). Using WordNet to disambiguate word senses for text retrieval. Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

VOORHEES, E. M. (1994). Query expansion using lexical-semantic relations. SIGIR'94, Springer.

VOORHEES, E. M. (1999). The TREC-8 Question Answering Track Report. TREC.

W3C (2004a). "Architecture of the World Wide Web, Volume One [Online]." Avaliable from http://www.w3.org/TR/2004/REC-webarch-20041215/. [Accessed 18 April 2011].

W3C (2004b). "OWL Web Ontology Language Guide [Online]." Avaliable from http://www.w3.org/TR/owl-guide/. [Accessed 15 April 2011].

W3C (2004c). "RDF Primer [Online]." Avaliable from http://www.w3.org/TR/rdf-primer/. [Accessed 1 May 2011].

W3C (2004d). "RDF Vocabulary Description Language 1.0: RDF Schema [Online]." Avaliable from http://www.w3.org/TR/2004/REC-rdf-schema-20040210/. [Accessed

2 May 2011].

XU, J.&CROFT, W. (1999). "Cluster-based language models for distributed retrieval [Online]." Avaliable from http://portal.acm.org/citation.cfm?id=312687. [Accessed 8 April 2010].

XU, J.&CROFT, W. B. (2000). "Improving the effectiveness of information retrieval with local context analysis." ACM Transactions on Information Systems (TOIS) 18(1), pp. 79-112.

XU, J., WEISCHEDEL, R., et al. (2001). "Evaluating a probabilistic model for cross-lingual information retrieval [Online]." Avaliable from http://portal.acm.org/citation.cfm?id=383968. [Accessed 8 July 2013].

XU, Q.-L., KUN, Y., et al. (2008). "A study on ontology-driven geospatial-information retrieval systems in the semantic web." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing 37, pp. 733-738.

ZARAGOZA, H., HIEMSTRA, D., et al. (2003). "Bayesian extension to the language model for ad hoc information retrieval [Online]." Avaliable from http://portal.acm.org/citation.cfm?id=860439. [Accessed 15 June 2013].

ZHAI, C. (2009). Statistical Language Models for Information Retrieval, Morgan & Claypool Publishers.

ZHAI, C.&LAFFERTY, J. (2001). Model-based feedback in the language modeling approach to information retrieval. Proceedings of the tenth international conference on Information and knowledge management, ACM.

ZHAI, C.&LAFFERTY, J. (2002). Two-stage language models for information retrieval. Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

ZHAI, C.&LAFFERTY, J. (2004). "A study of smoothing methods for language models applied to information retrieval." ACM Trans. Inf. Syst. 22(2), pp. 179-214.

ZHANG, Y., CALLAN, J., et al. (2002). "Novelty and redundancy detection in adaptive filtering [Online]." Avaliable from http://portal.acm.org/citation.cfm?id=564376.564393. [Accessed 1 June 2013].

ZHOU, X. (2008a). Semantics-based Language Models for Information Retrieval and Text Mining, Drexel University. PHD.

ZHOU, X. (2008b). Semantics-based Language Models for Information Retrieval and Text Mining. Computer science, Drexel University.

APPENDICES

Appendix A: Glossary

URI/IRI A URI (Uniform Resource Identifier) provide a mechanism to identify or name unique reference to entities and relations. IRI is an acronym for Internationalized Resource Identifier, which is a special case for URI by using characters beyond ASCII.

XML is abbreviation for Extensible Mark-up Language. It offers a standard to compose information so that it can be more easily shared. It is a textual data format with strong support via Unicode.

Namespaces are elementary to XML, which enable differentiation between combinations of documents.

XML Query is a query and functional standard language used for query collection of XML data.

XML Schema is used to mechanism to control the structure of XML documents.

Logic and Proof as a reasoning system, they provided top of the ontology layer to make new inference. According to the new inference result, the computer can make deduction or further manipulation to satisfy user requirements.

Trust policy is a subjective procedure used for evaluating the trustworthiness of the information on the web, in order to provide an assurance of its quality.

Appendix B: Ontology contexts

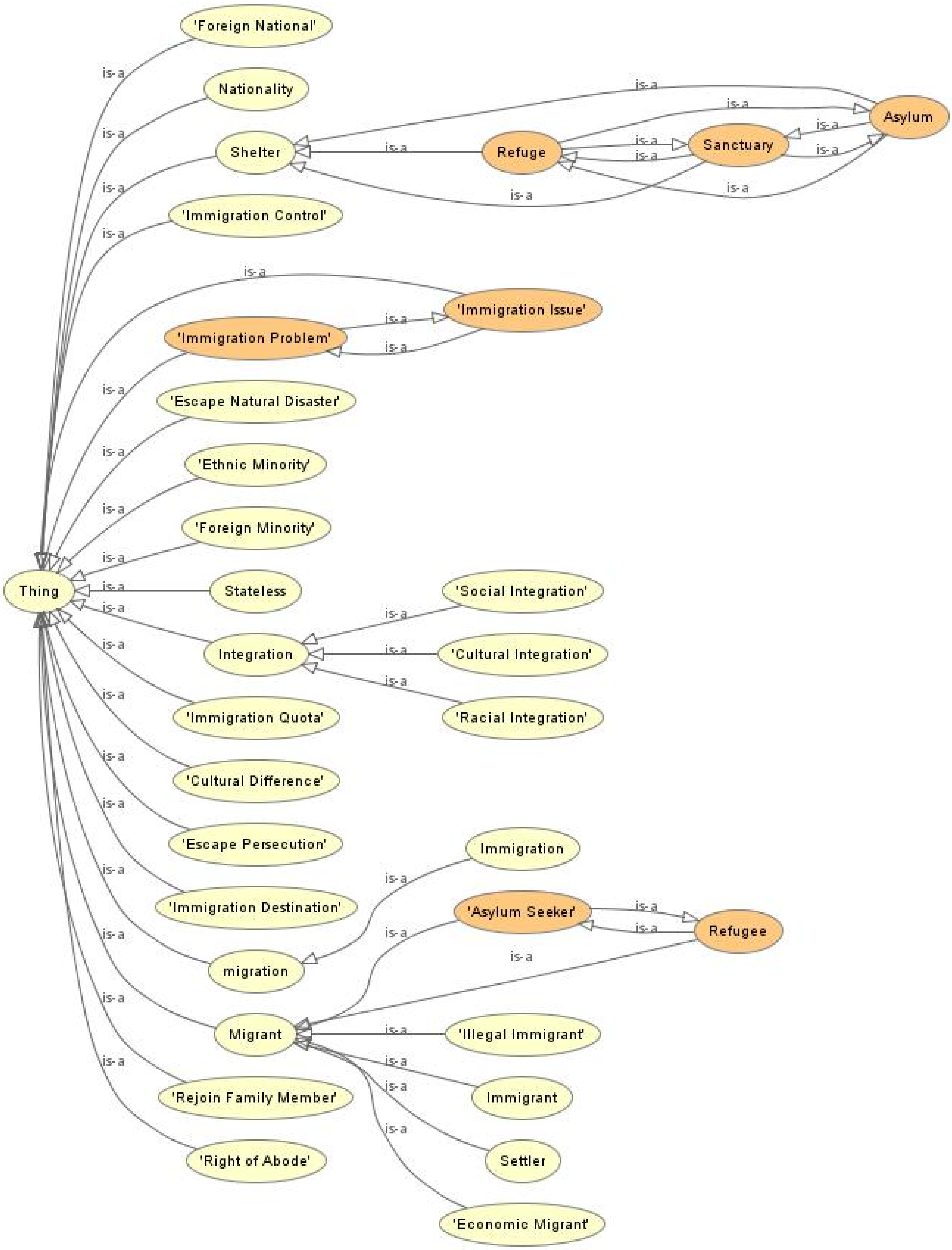
1. T401 Immigration

2. T403 Osteoporosis

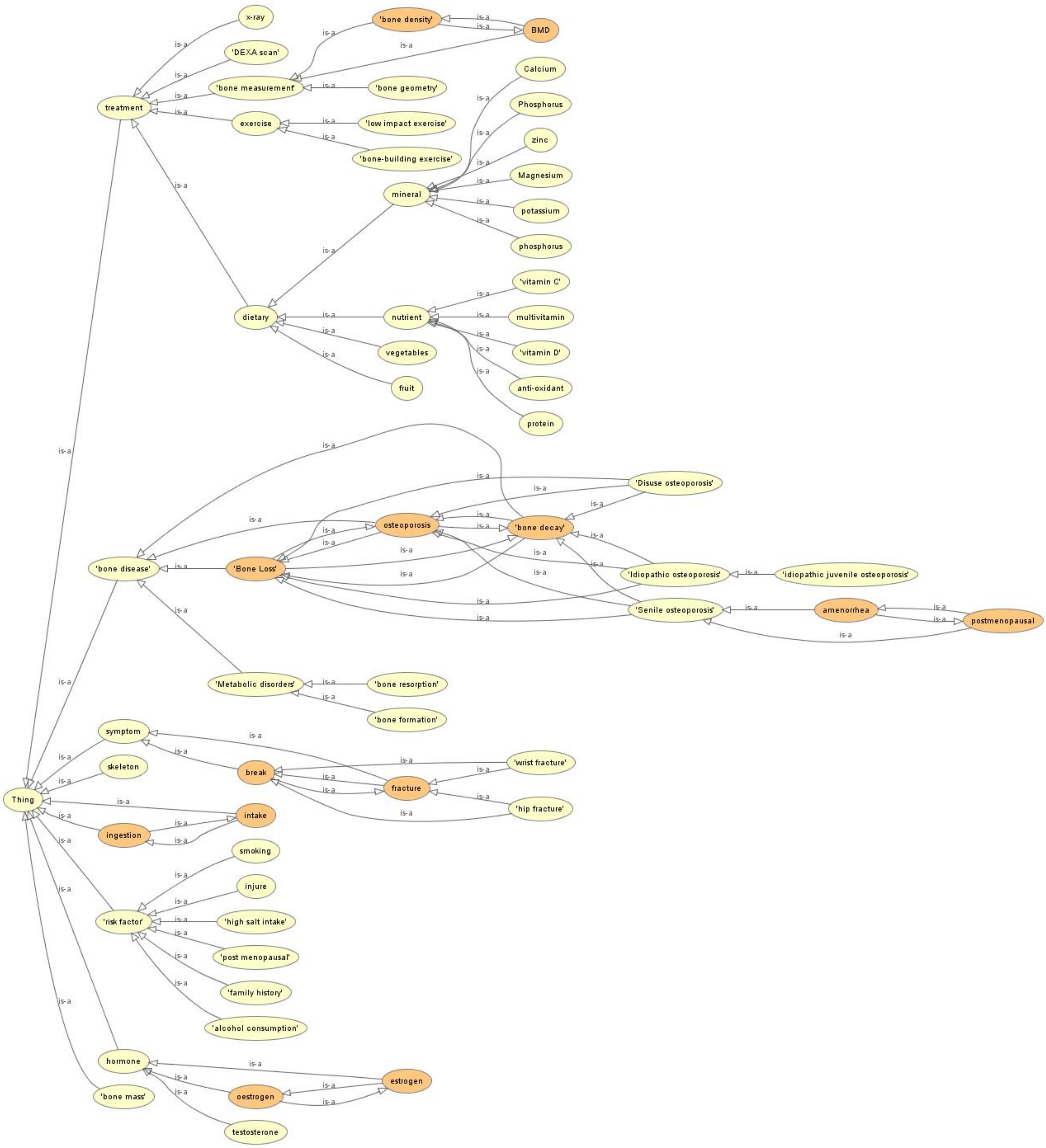
3. T416 Hydro-electric

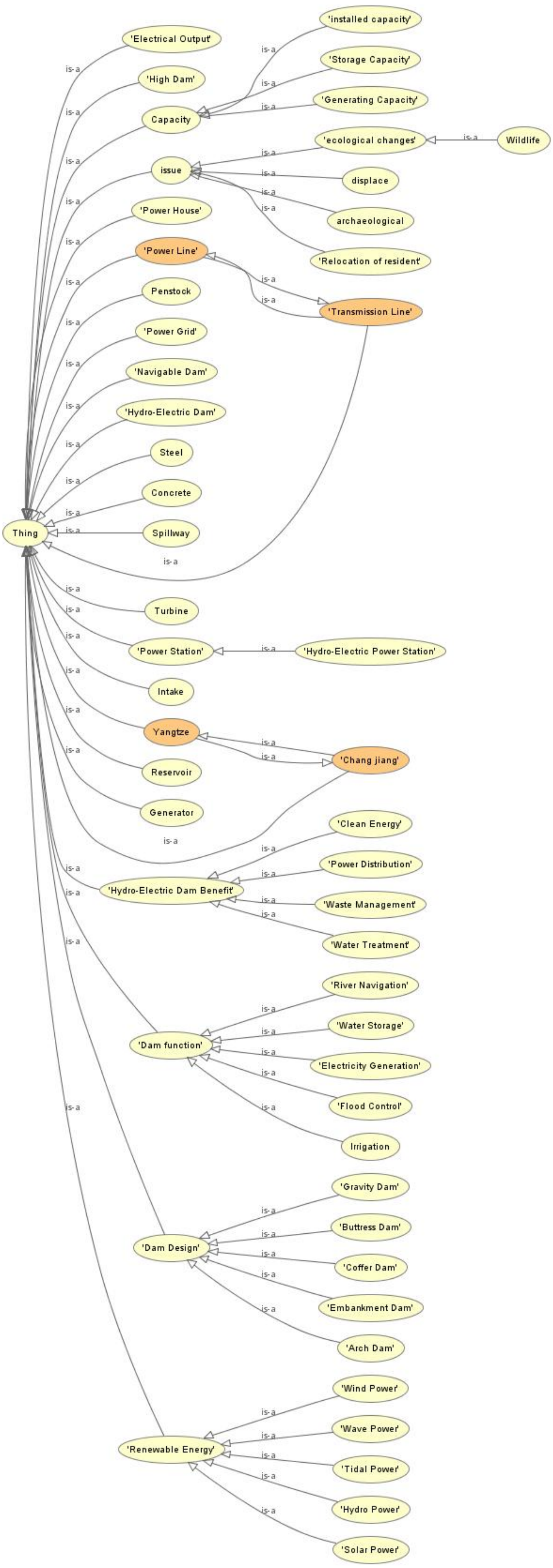
4. T403 Robotic

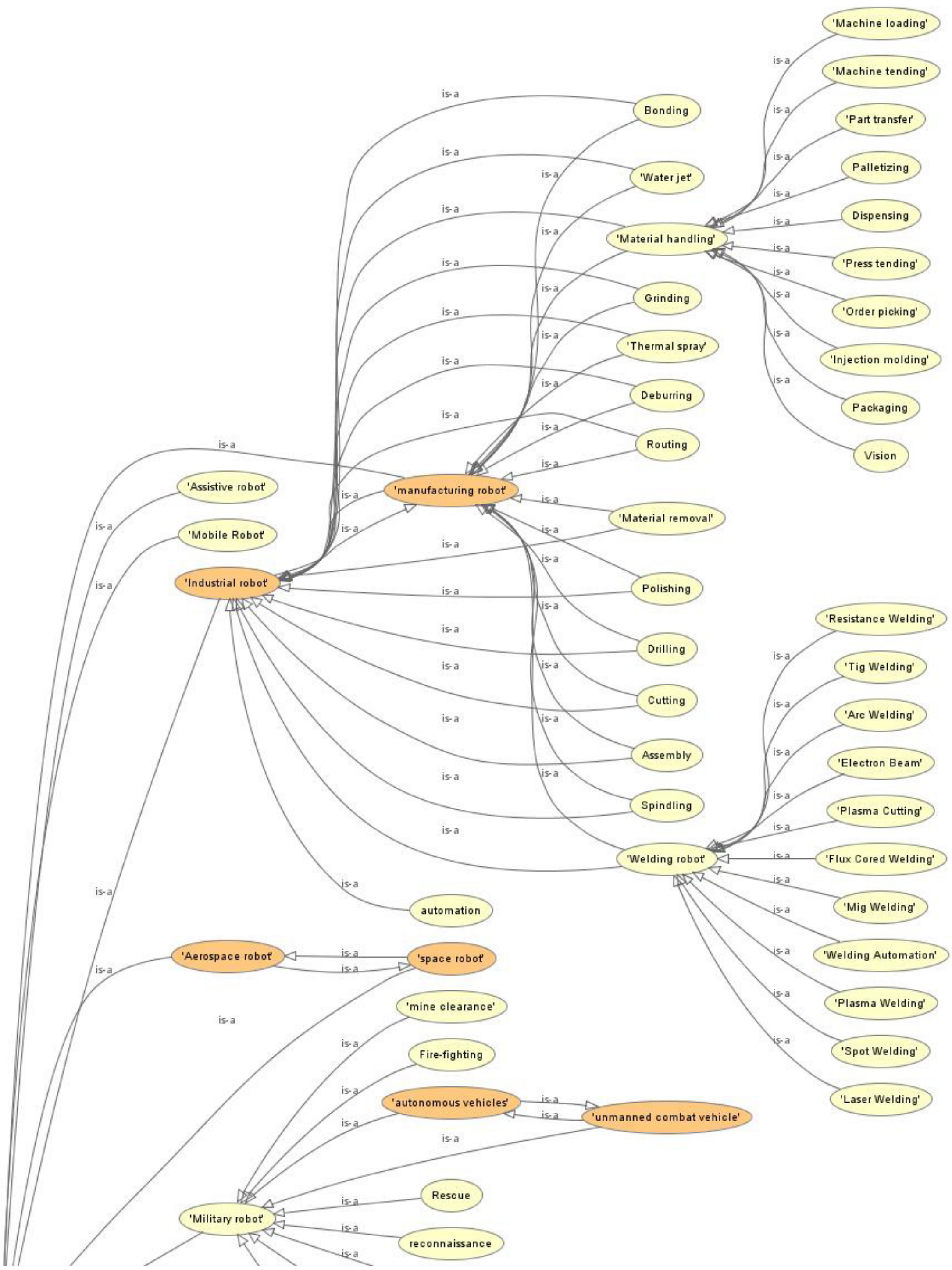
5. T438 Tourism

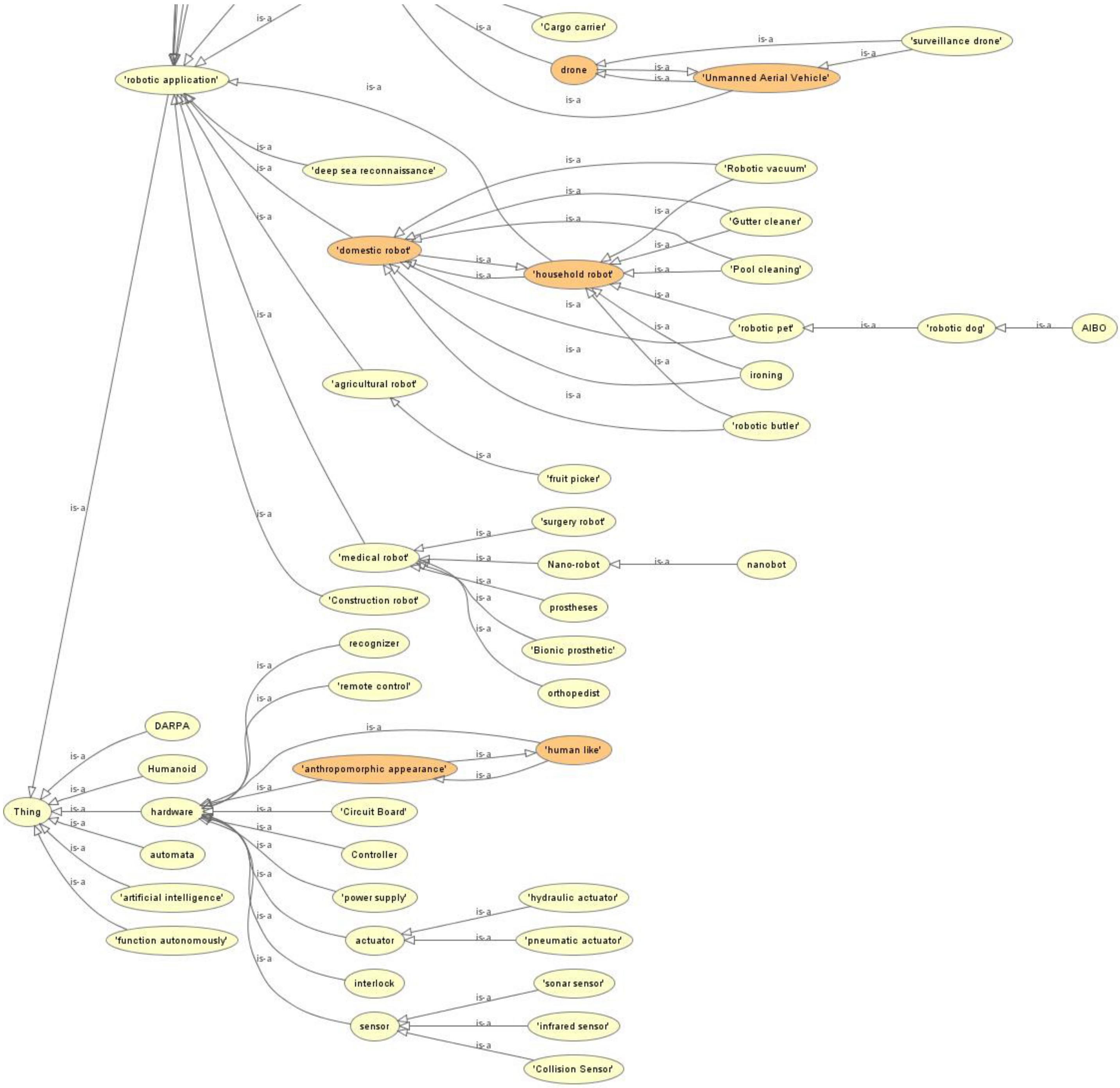


'Foreign National'
Nationality
Shelter
Refuge
Sanctuary
Asylum
'Immigration Control'
'Immigration Issue'
'Immigration Problem'
'Escape Natural Disaster'
'Ethnic Minority'
'Foreign Minority'
Thing
Stateless
'Social Integration'
Integration
'Cultural Integration'
'Racial Integration'
'Immigration Quota'
'Cultural Difference'
'Escape Persecution'
Immigration
'Immigration Destination'
'Asylum Seeker'
Refugee
migration
Migrant
'Illegal Immigrant'
Immigrant
Settler
'Rejoin Family Member'
'Right of Abode'
'Economic Migrant'
is-a

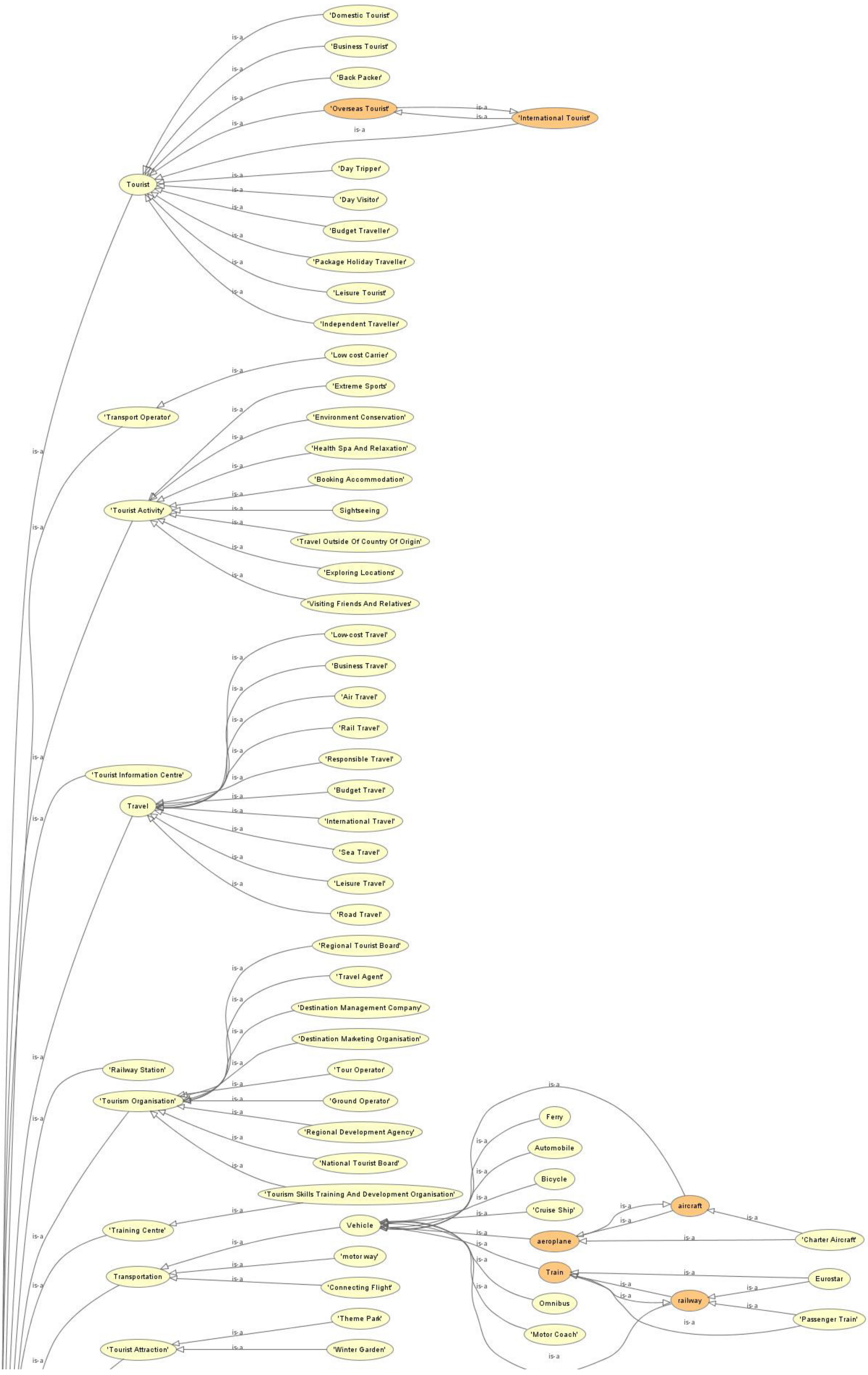

Thing
treatment
x-ray
'DEXA scan'
'bone measurement'
exercise
'bone density'
BMD
'bone geometry'
'low impact exercise'
'bone-building exercise'
mineral
Calcium
Phosphorus
zinc
Magnesium
potassium
phosphorus
dietary
nutrient
vegetables
fruit
'vitamin C'
multivitamin
'vitamin D'
anti-oxidant
protein
'bone disease'
'Bone Loss'
osteoporosis
'bone decay'
'Disuse osteoporosis'
'Idiopathic osteoporosis'
'idiopathic juvenile osteoporosis'
'Senile osteoporosis'
amenorrhea
postmenopausal
'Metabolic disorders'
'bone resorption'
'bone formation'
symptom
skeleton
break
fracture
'wrist fracture'
'hip fracture'
intake
ingestion
'risk factor'
smoking
injure
'high salt intake'
'post menopausal'
'family history'
'alcohol consumption'
hormone
estrogen
oestrogen
testosterone
'bone mass'
is-a

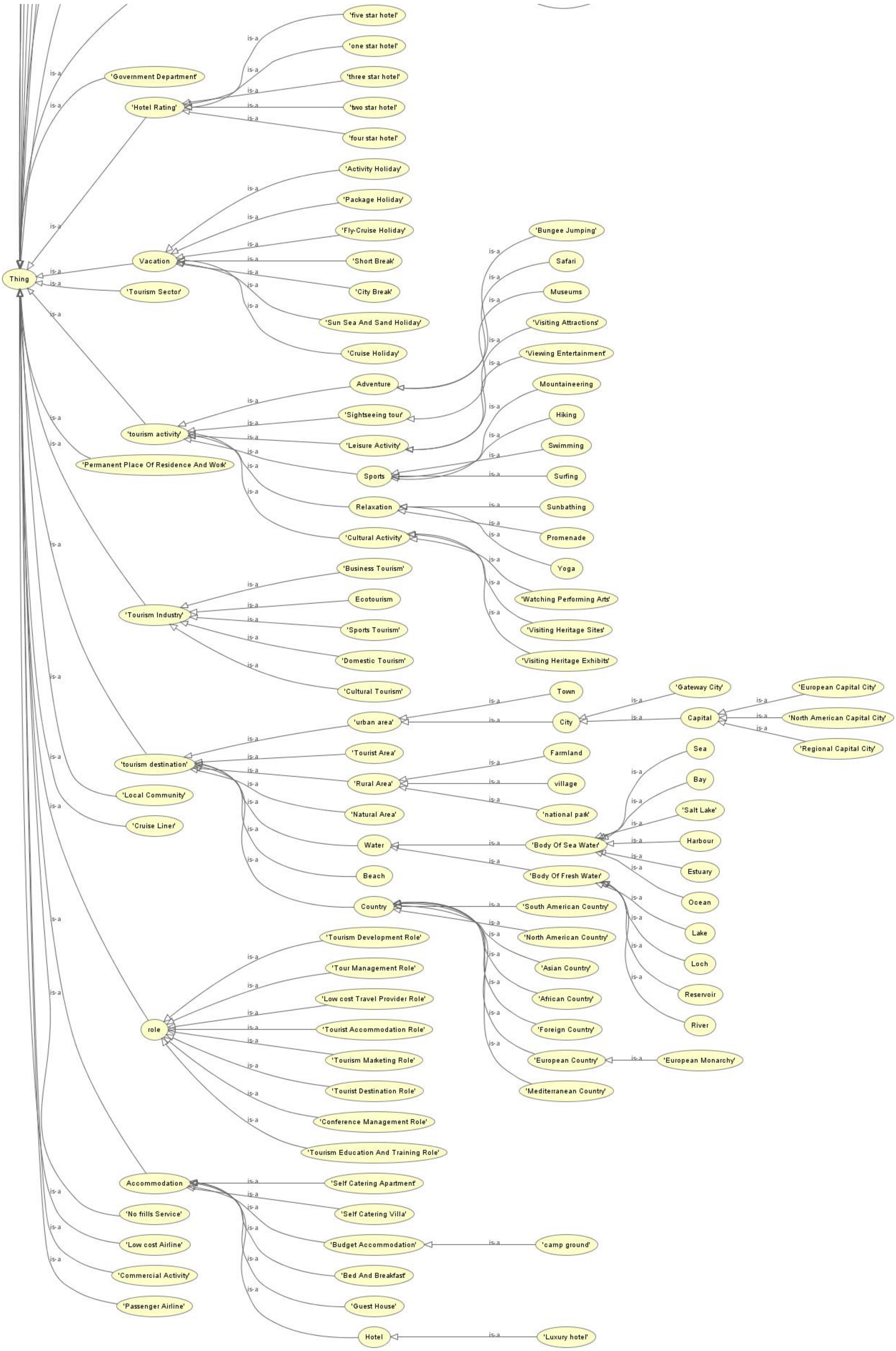

Thing
'Electrical Output'
'High Dam'
Capacity
'installed capacity'
'Storage Capacity'
'Generating Capacity'
issue
'ecological changes'
Wildlife
displace
archaeological
'Relocation of resident'
'Power House'
'Power Line'
'Transmission Line'
Penstock
'Power Grid'
'Navigable Dam'
'Hydro-Electric Dam'
Steel
Concrete
Spillway
Turbine
'Power Station'
'Hydro-Electric Power Station'
Intake
Yangtze
'Chang jiang'
Reservoir
Generator
'Hydro-Electric Dam Benefit'
'Clean Energy'
'Power Distribution'
'Waste Management'
'Water Treatment'
'Dam function'
'River Navigation'
'Water Storage'
'Electricity Generation'
'Flood Control'
Irrigation
'Dam Design'
'Gravity Dam'
'Buttress Dam'
'Coffer Dam'
'Embankment Dam'
'Arch Dam'
'Renewable Energy'
'Wind Power'
'Wave Power'
'Tidal Power'
'Hydro Power'
'Solar Power'
is-a


'Machine loading'
'Machine tending'
'Part transfer'
Palletizing
Dispensing
'Press tending'
'Order picking'
'Injection molding'
Packaging
Vision
Bonding
'Water jet'
'Material handling'
Grinding
'Thermal spray'
Deburring
Routing
'manufacturing robot'
'Material removal'
Polishing
Drilling
Cutting
Assembly
Spindling
'Welding robot'
'Resistance Welding'
'Tig Welding'
'Arc Welding'
'Electron Beam'
'Plasma Cutting'
'Flux Cored Welding'
'Mig Welding'
'Welding Automation'
'Plasma Welding'
'Spot Welding'
'Laser Welding'
'Assistive robot'
'Mobile Robot'
'Industrial robot'
automation
'Aerospace robot'
'space robot'
'mine clearance'
Fire-fighting
'autonomous vehicles'
'unmanned combat vehicle'
Rescue
'Military robot'
reconnaissance
is-a



'robotic application'
'Cargo carrier'
drone
'surveillance drone'
'Unmanned Aerial Vehicle'
'deep sea reconnaissance'
'Robotic vacuum'
'Gutter cleaner'
'domestic robot'
'household robot'
'Pool cleaning'
'robotic pet'
'robotic dog'
AIBO
ironing
'robotic butler'
'agricultural robot'
'fruit picker'
'surgery robot'
'medical robot'
Nano-robot
nanobot
prostheses
'Construction robot'
'Bionic prosthetic'
orthopedist
recognizer
'remote control'
DARPA
'human like'
Humanoid
'anthropomorphic appearance'
Thing
hardware
'Circuit Board'
automata
Controller
'artificial intelligence'
'power supply'
'hydraulic actuator'
'function autonomously'
actuator
'pneumatic actuator'
interlock
'sonar sensor'
sensor
'infrared sensor'
'Collision Sensor'
is-a


'Domestic Tourist'
'Business Tourist'
'Back Packer'
'Overseas Tourist'
'International Tourist'
Tourist
'Day Tripper'
'Day Visitor'
'Budget Traveller'
'Package Holiday Traveller'
'Leisure Tourist'
'Independent Traveller'
'Low cost Carrier'
'Transport Operator'
'Extreme Sports'
'Environment Conservation'
'Health Spa And Relaxation'
'Booking Accommodation'
'Tourist Activity'
Sightseeing
'Travel Outside Of Country Of Origin'
'Exploring Locations'
'Visiting Friends And Relatives'
'Low-cost Travel'
'Business Travel'
'Air Travel'
'Rail Travel'
'Responsible Travel'
'Tourist Information Centre'
Travel
'Budget Travel'
'International Travel'
'Sea Travel'
'Leisure Travel'
'Road Travel'
'Regional Tourist Board'
'Travel Agent'
'Destination Management Company'
'Destination Marketing Organisation'
'Railway Station'
'Tour Operator'
'Tourism Organisation'
'Ground Operator'
'Regional Development Agency'
'National Tourist Board'
'Tourism Skills Training And Development Organisation'
Ferry
Automobile
Bicycle
'Cruise Ship'
aircraft
'Training Centre'
Vehicle
aeroplane
'Charter Aircraft'
'motor way'
Transportation
'Connecting Flight'
Train
Eurostar
railway
Omnibus
'Passenger Train'
'Theme Park'
'Tourist Attraction'
'Motor Coach'
'Winter Garden'
is-a


Thing
'Government Department'
'Hotel Rating'
'five star hotel'
'one star hotel'
'three star hotel'
'two star hotel'
'four star hotel'
Vacation
'Activity Holiday'
'Package Holiday'
'Fly-Cruise Holiday'
'Short Break'
'City Break'
'Sun Sea And Sand Holiday'
'Cruise Holiday'
'Tourism Sector'
'tourism activity'
Adventure
'Bungee Jumping'
Safari
Museums
'Visiting Attractions'
'Viewing Entertainment'
'Sightseeing tour'
'Leisure Activity'
Sports
Mountaineering
Hiking
Swimming
Surfing
Relaxation
Sunbathing
Promenade
Yoga
'Cultural Activity'
'Watching Performing Arts'
'Visiting Heritage Sites'
'Visiting Heritage Exhibits'
'Permanent Place Of Residence And Work'
'Tourism Industry'
'Business Tourism'
Ecotourism
'Sports Tourism'
'Domestic Tourism'
'Cultural Tourism'
'tourism destination'
'urban area'
Town
City
'Gateway City'
Capital
'European Capital City'
'North American Capital City'
'Regional Capital City'
'Tourist Area'
'Rural Area'
Farmland
village
'national park'
'Natural Area'
Water
'Body Of Sea Water'
'Body Of Fresh Water'
Sea
Bay
'Salt Lake'
Harbour
Estuary
Ocean
Lake
Loch
Reservoir
River
Beach
Country
'South American Country'
'North American Country'
'Asian Country'
'African Country'
'Foreign Country'
'European Country'
'European Monarchy'
'Mediterranean Country'
'Local Community'
'Cruise Liner'
role
'Tourism Development Role'
'Tour Management Role'
'Low cost Travel Provider Role'
'Tourist Accommodation Role'
'Tourism Marketing Role'
'Tourist Destination Role'
'Conference Management Role'
'Tourism Education And Training Role'
Accommodation
'Self Catering Apartment'
'Self Catering Villa'
'Budget Accommodation'
'camp ground'
'Bed And Breakfast'
'Guest House'
Hotel
'Luxury hotel'
'No frills Service'
'Low cost Airline'
'Commercial Activity'
'Passenger Airline'
is-a

Appendix C: Precision and Recall Data

T401 Query 1

T401 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	41.67%	38.46%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	33.33%	33.33%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	18.92%	22.58%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	19.15%	15.52%	81.82%	81.82%	78.26%	78.26%	78.26%	81.82%	81.82%	81.82%	78.26%	78.26%	78.26%	81.82%	81.82%	81.82%	78.26%	78.26%
50%	19.01%	13.30%	79.31%	79.31%	76.67%	76.67%	76.67%	79.31%	79.31%	79.31%	76.67%	76.67%	76.67%	79.31%	79.31%	79.31%	76.67%	76.67%
60%	13.57%	8.91%	67.50%	67.50%	72.97%	45%	49.09%	67.50%	67.50%	67.50%	72.97%	45%	49.09%	67.50%	67.50%	67.50%	72.97%	45%
70%	3.56%	2.02%	24.43%	24.43%	26.23%	16.84%	20.65%	24.81%	24.43%	24.43%	26.23%	16.84%	20.65%	24.81%	24.43%	24.43%	26.23%	16.84%
80%	0.59%	0.71%	5.62%	5.62%	7.38%	4.73%	4.54%	6.05%	5.62%	5.62%	7.38%	4.73%	4.54%	6.05%	5.62%	5.62%	7.38%	4.73%
90%	0.59%	0.57%	0.78%	0.78%	0.83%	0.88%	0.67%	0.77%	0.76%	0.76%	0.83%	0.88%	0.67%	0.77%	0.76%	0.76%	0.83%	0.88%
100%	0.35%	0.15%	0.43%	0.43%	0.15%	0.14%	0.40%	0.42%	0.41%	0.41%	0.15%	0.14%	0.40%	0.42%	0.41%	0.41%	0.15%	0.14%
Avg30	31.31%	31.46%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%
AVG100	15.07%	13.55%	51.25%	51.25%	51.51%	47.52%	48.29%	51.33%	51.25%	51.25%	51.51%	47.52%	48.29%	51.33%	51.25%	51.25%	51.51%	47.52%

T401 Query 2

T401 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
50%	85.19%	85.19%	79.31%	85.19%	85.19%	85.19%	85.19%	79.31%	79.31%	85.19%	85.19%	85.19%	85.19%	79.31%	79.31%	85.19%	85.19%	85.19%
60%	84.38%	81.82%	75%	75%	75%	77.14%	77.14%	75%	75%	75%	75%	77.14%	77.14%	75%	75%	75%	75%	77.14%
70%	71.11%	74.42%	52.46%	52.46%	50%	42.11%	46.38%	52.46%	52.46%	52.46%	50%	42.11%	46.38%	52.46%	52.46%	52.46%	50%	42.11%
80%	18.18%	69.23%	41.38%	41.38%	34.95%	38.30%	39.56%	41.38%	41.38%	41.38%	34.95%	38.30%	39.13%	41.38%	41.38%	41.38%	34.95%	38.30%
90%	1.75%	8.09%	4.58%	4.67%	3.57%	2.44%	3.77%	4.15%	4.15%	4.33%	3.51%	2.25%	4.06%	4.24%	4.24%	4.33%	3.51%	2.25%
100%	0.93%	0.80%	0.78%	0.73%	0.70%	0.56%	0.64%	0.67%	0.67%	0.64%	0.61%	0.48%	0.59%	0.63%	0.63%	0.63%	0.61%	0.48%
Avg30	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%	84.22%
AVG100	60.42%	66.22%	59.62%	60.21%	59.21%	58.84%	59.53%	59.56%	59.56%	60.16%	59.19%	58.81%	59.51%	59.57%	59.57%	60.16%	59.19%	58.81%

T401 Query 3

T401 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	90%	90%	90%	85.71%	85.71%	85.71%	90%	90%	90%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	88.46%	88.46%	88.46%	88.46%	79.31%	88.46%	88.46%	88.46%	88.46%	88.46%	79.31%	88.46%	88.46%	88.46%	88.46%	88.46%	79.31%	88.46%
60%	84.38%	84.38%	84.38%	81.82%	81.82%	81.82%	84.38%	84.38%	84.38%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
70%	82.05%	82.05%	82.05%	82.05%	76.19%	82.05%	82.05%	82.05%	82.05%	82.05%	76.19%	82.05%	82.05%	82.05%	82.05%	82.05%	76.19%	82.05%
80%	69.23%	69.23%	69.23%	72%	75%	73.47%	69.23%	69.23%	69.23%	72%	75%	73.47%	72%	72%	72%	72%	75%	73.47%
90%	0.56%	0.64%	0.64%	0.65%	0.75%	0.68%	0.56%	0.64%	0.64%	0.65%	0.75%	0.68%	0.57%	0.65%	0.65%	0.65%	0.75%	0.68%
100%	0.40%	0.42%	0.42%	0.43%	0.45%	0.20%	0.40%	0.42%	0.42%	0.43%	0.45%	0.20%	0.40%	0.43%	0.43%	0.42%	0.45%	0.20%
Avg30	94.44%	94.44%	94.44%	92.50%	92.50%	92.50%	94.44%	94.44%	94.44%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%
AVG100	69.84%	69.85%	69.85%	68.86%	67.67%	68.99%	69.84%	69.85%	69.85%	68.86%	67.67%	68.99%	68.85%	68.86%	68.86%	68.86%	67.67%	68.99%

T401 Query 4

T401 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	83.33%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	81.82%	100%	100%	90%	90%	90%	100%	100%	100%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	87.50%	82.35%	82.35%	87.50%	87.50%	87.50%	82.35%	82.35%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	85.71%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
50%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%
60%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%	87.10%
70%	82.05%	82.05%	82.05%	84.21%	84.21%	84.21%	82.05%	82.05%	82.05%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	29.75%	63.16%	63.16%	63.16%	62.07%	62.07%	63.16%	63.16%	63.16%	63.16%	62.07%	62.07%	63.16%	63.16%	63.16%	63.16%	62.07%	62.07%
90%	2.01%	4.37%	4.37%	7.24%	4.66%	1.20%	4.08%	4.37%	4.37%	7.24%	4.66%	1.19%	6.17%	6.80%	6.80%	7.24%	4.66%	1.19%
100%	0.59%	0.58%	0.58%	0.49%	0.46%	0.40%	0.53%	0.55%	0.55%	0.48%	0.43%	0.36%	0.45%	0.47%	0.47%	0.47%	0.42%	0.32%
Avg30	84.22%	94.12%	94.12%	92.50%	92.50%	92.50%	94.12%	94.12%	94.12%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%
AVG100	62.50%	68.66%	68.66%	68.67%	68.30%	67.95%	68.63%	68.66%	68.66%	68.67%	68.30%	67.94%	68.56%	68.62%	68.62%	68.67%	68.30%	67.94%

T401 Query 5

T401 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%	82.35%
40%	81.82%	78.26%	81.82%	81.82%	75%	81.82%	85.71%	81.82%	81.82%	81.82%	75%	81.82%	85.71%	81.82%	81.82%	81.82%	75%	81.82%
50%	85.19%	71.88%	79.31%	79.31%	79.31%	74.19%	88.46%	79.31%	79.31%	79.31%	79.31%	74.19%	88.46%	79.31%	79.31%	79.31%	79.31%	74.19%
60%	87.10%	72.97%	77.14%	77.14%	81.82%	77.14%	84.38%	77.14%	77.14%	77.14%	81.82%	77.14%	84.38%	77.14%	77.14%	77.14%	81.82%	77.14%
70%	72.73%	66.67%	65.31%	65.31%	68.09%	74.42%	61.54%	65.31%	65.31%	65.31%	68.09%	74.42%	61.54%	65.31%	65.31%	65.31%	68.09%	74.42%
80%	13.38%	30.25%	31.03%	31.03%	22.22%	24.32%	21.69%	31.03%	31.03%	31.03%	22.22%	24.32%	21.69%	31.03%	31.03%	31.03%	22.22%	24.32%
90%	0.97%	1.47%	5.37%	5.37%	4.08%	1.11%	3.25%	5.21%	5.37%	5.37%	4.08%	1.11%	3.25%	5.21%	5.37%	5.37%	4.08%	1.11%
100%	0.51%	0.82%	0.76%	0.76%	0.64%	0.63%	0.60%	0.75%	0.75%	0.75%	0.64%	0.62%	0.60%	0.75%	0.75%	0.75%	0.64%	0.62%
Avg30	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%	82.50%
AVG100	58.92%	56.98%	58.82%	58.82%	57.87%	58.11%	59.31%	58.81%	58.82%	58.82%	57.87%	58.11%	59.31%	58.81%	58.82%	58.82%	57.87%	58.11%

T401 Query 6

T401 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	62.07%	62.07%	62.07%	85.71%	85.71%	85.71%	60%	62.07%	62.07%	85.71%	85.71%	85.71%	85.71%	64.29%	64.29%	85.71%	85.71%	85.71%
50%	65.71%	65.71%	65.71%	62.16%	60.53%	79.31%	65.71%	65.71%	65.71%	62.16%	60.53%	79.31%	62.16%	62.16%	62.16%	62.16%	60.53%	79.31%
60%	67.50%	67.50%	67.50%	56.25%	64.29%	67.50%	69.23%	67.50%	67.50%	56.25%	64.29%	67.50%	56.25%	56.25%	56.25%	56.25%	64.29%	67.50%
70%	68.09%	68.09%	68.09%	58.18%	68.09%	71.11%	69.57%	68.09%	68.09%	58.18%	68.09%	71.11%	57.14%	57.14%	57.14%	58.18%	68.09%	71.11%
80%	23.08%	25.35%	25.35%	36%	31.03%	35.64%	22.64%	25.35%	25.35%	33.33%	30.77%	35.64%	37.90%	35.64%	35.64%	33.33%	30.77%	35.64%
90%	1.43%	1.71%	1.71%	9.81%	4.51%	1.57%	1.42%	1.69%	1.69%	9.62%	4.47%	1.55%	5.08%	8.25%	8.25%	9.62%	4.47%	1.55%
100%	0.83%	0.90%	0.90%	1.15%	1.07%	0.90%	0.82%	0.90%	0.90%	1.15%	1.06%	0.89%	1.03%	1.06%	1.06%	1.08%	1.04%	0.88%
Avg30	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%
AVG100	57.20%	57.47%	57.47%	59.26%	59.86%	62.51%	57.27%	57.46%	57.46%	58.97%	59.82%	62.51%	58.86%	56.81%	56.81%	58.97%	59.82%	62.50%

T401 Query 7

T401 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	83.33%	83.33%	83.33%	100%	100%	100%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	34.62%	34.62%	34.62%	90%	90%	90%	34.62%	34.62%	34.62%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	18.18%	16.67%	16.67%	93.33%	93.33%	93.33%	18.18%	16.67%	16.67%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	15.52%	12.77%	12.77%	85.71%	81.82%	81.82%	15.52%	12.77%	12.77%	85.71%	81.82%	81.82%	85.71%	85.71%	85.71%	85.71%	81.82%	81.82%
50%	11.86%	12.43%	12.43%	74.19%	76.67%	76.67%	11.86%	12.43%	12.43%	74.19%	76.67%	76.67%	76.67%	76.67%	76.67%	74.19%	76.67%	76.67%
60%	10.39%	9.51%	9.51%	71.05%	71.05%	71.05%	10.39%	9.51%	9.51%	71.05%	71.05%	71.05%	69.23%	71.05%	71.05%	71.05%	71.05%	71.05%
70%	7.79%	8.82%	8.82%	44.44%	32.65%	36.36%	7.79%	8.82%	8.82%	44.44%	32.65%	36.36%	38.10%	45.07%	45.07%	44.44%	32.65%	36.36%
80%	1.73%	1.87%	1.87%	27.48%	18.46%	22.93%	1.73%	1.87%	1.87%	27.48%	18.46%	22.93%	24.49%	30.77%	30.77%	27.48%	18.46%	22.93%
90%	0.66%	0.77%	0.77%	0.73%	0.77%	0.84%	0.66%	0.77%	0.77%	0.73%	0.77%	0.84%	0.63%	0.73%	0.73%	0.71%	0.77%	0.82%
100%	0.40%	0.43%	0.43%	0.48%	0.50%	0.27%	0.40%	0.43%	0.43%	0.48%	0.50%	0.27%	0.43%	0.48%	0.48%	0.46%	0.50%	0.26%
Avg30	50.93%	50.43%	50.43%	88.89%	88.89%	88.89%	50.93%	50.43%	50.43%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%
AVG100	20.11%	19.79%	19.79%	57.08%	54.86%	55.66%	20.11%	19.79%	19.79%	57.08%	54.86%	55.66%	56.19%	57.72%	57.72%	57.07%	54.86%	55.66%

T401 Query 8

T401 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	100%	100%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	82.35%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	78.26%	62.07%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	63.89%	45.10%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%	88.46%
60%	48.21%	46.55%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
70%	26.45%	36.36%	80%	80%	76.19%	76.19%	76.19%	80%	80%	80%	76.19%	76.19%	76.19%	80%	80%	80%	76.19%	76.19%
80%	20.93%	19.15%	40.91%	40.91%	48.65%	29.51%	34.95%	40.91%	40.91%	40.91%	48.65%	29.51%	34.95%	40.91%	40.91%	40.91%	48.65%	29.51%
90%	2.87%	2.37%	3.01%	3.01%	2.16%	1.17%	2.11%	2.84%	3.01%	3.01%	2.16%	1.17%	2.11%	2.84%	3.01%	3.01%	2.16%	1.17%
100%	0.62%	0.72%	0.67%	0.67%	0.76%	0.42%	0.61%	0.67%	0.65%	0.65%	0.75%	0.41%	0.61%	0.67%	0.65%	0.65%	0.75%	0.41%
Avg30	94.12%	97.78%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%
AVG100	52.36%	50.57%	67.21%	67.21%	67.53%	65.48%	66.14%	67.19%	67.21%	67.21%	67.53%	65.48%	66.14%	67.19%	67.21%	67.21%	67.53%	65.48%

T401 Query 9

T401 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	75%	81.82%	81.82%	100%	100%	100%	81.82%	81.82%	81.82%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30%	82.35%	87.50%	87.50%	77.78%	77.78%	77.78%	87.50%	87.50%	87.50%	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%	77.78%
40%	81.82%	85.71%	85.71%	78.26%	78.26%	78.26%	85.71%	85.71%	85.71%	78.26%	78.26%	78.26%	78.26%	78.26%	78.26%	78.26%	78.26%	78.26%
50%	82.14%	85.19%	85.19%	82.14%	82.14%	82.14%	88.46%	85.19%	85.19%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%	82.14%
60%	84.38%	87.10%	87.10%	84.38%	84.38%	84.38%	87.10%	87.10%	87.10%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%
70%	78.05%	82.05%	82.05%	84.21%	84.21%	84.21%	82.05%	82.05%	82.05%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	27.27%	41.86%	41.86%	73.47%	73.47%	73.47%	50.70%	41.86%	41.86%	73.47%	73.47%	73.47%	73.47%	73.47%	73.47%	73.47%	73.47%	73.47%
90%	3.64%	2.02%	2.02%	1.62%	0.72%	0.49%	1.63%	2.02%	2.02%	1.61%	0.66%	0.46%	1.24%	1.48%	1.48%	1.59%	0.60%	0.35%
100%	1.08%	0.89%	0.89%	0.87%	0.52%	0.47%	0.83%	0.86%	0.86%	0.85%	0.49%	0.43%	0.75%	0.81%	0.81%	0.80%	0.38%	0.33%
Avg30	85.78%	89.77%	89.77%	92.59%	92.59%	92.59%	89.77%	89.77%	89.77%	92.59%	92.59%	92.59%	92.59%	92.59%	92.59%	92.59%	92.59%	92.59%
AVG100	61.57%	65.41%	65.41%	68.27%	68.15%	68.12%	66.58%	65.41%	65.41%	68.27%	68.14%	68.11%	68.22%	68.25%	68.25%	68.26%	68.12%	68.09%

T401 Query 10

T401 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	90%	90%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	92%	92%	88.46%	88.46%	82.14%	82.14%	85.19%	88.46%	88.46%	88.46%	82.14%	82.14%	85.19%	88.46%	88.46%	88.46%	82.14%	82.14%
60%	90%	87.10%	81.82%	81.82%	81.82%	79.41%	81.82%	81.82%	81.82%	81.82%	81.82%	79.41%	81.82%	81.82%	81.82%	81.82%	81.82%	79.41%
70%	80%	82.05%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%	72.73%
80%	69.23%	72%	63.16%	63.16%	43.37%	42.35%	53.73%	63.16%	63.16%	63.16%	43.37%	42.35%	53.73%	63.16%	63.16%	63.16%	43.37%	42.35%
90%	0.70%	0.82%	3.17%	3.17%	2.36%	0.74%	2.01%	3.29%	3.17%	3.17%	2.36%	0.74%	2.01%	3.29%	3.17%	3.17%	2.36%	0.74%
100%	0.45%	0.48%	0.55%	0.55%	0.58%	0.34%	0.51%	0.54%	0.54%	0.54%	0.58%	0.33%	0.51%	0.54%	0.54%	0.54%	0.58%	0.33%
Avg30	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%	86.94%
AVG100	68.32%	68.53%	65.64%	65.64%	62.95%	62.43%	64.25%	65.65%	65.64%	65.64%	62.95%	62.43%	64.25%	65.65%	65.64%	65.64%	62.95%	62.43%

T401 Query 1 entropy

T401 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	38.46%	27.78%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	34.62%	34.62%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	28.57%	36.84%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%
40%	15.93%	24.32%	81.82%	81.82%	81.82%	85.71%	85.71%	81.82%	81.82%	81.82%	81.82%	85.71%	85.71%	81.82%	81.82%	81.82%	81.82%	85.71%
50%	16.08%	20.54%	82.14%	82.14%	76.67%	79.31%	85.19%	82.14%	82.14%	82.14%	76.67%	79.31%	85.19%	82.14%	82.14%	82.14%	76.67%	79.31%
60%	10.47%	12.50%	72.97%	72.97%	67.50%	75%	75%	72.97%	72.97%	72.97%	67.50%	75%	75%	72.97%	72.97%	72.97%	67.50%	75%
70%	3.23%	7.92%	44.44%	44.44%	36.36%	28.57%	31.68%	46.38%	44.44%	44.44%	36.36%	28.57%	31.68%	46.38%	44.44%	44.44%	36.36%	28.57%
80%	0.73%	0.83%	11.15%	11.15%	14.63%	7.79%	8.55%	12.16%	11.15%	11.15%	14.63%	7.79%	8.55%	12.16%	11.15%	11.15%	14.63%	7.79%
90%	0.68%	0.69%	1.09%	1.09%	1.19%	1.24%	0.84%	1.04%	1.05%	1.05%	1.19%	1.21%	0.84%	1.04%	1.05%	1.05%	1.19%	1.21%
100%	0.30%	0.45%	0.48%	0.48%	0.15%	0.14%	0.43%	0.47%	0.45%	0.45%	0.15%	0.13%	0.43%	0.47%	0.45%	0.45%	0.15%	0.13%
Avg30	33.88%	33.08%	84.22%	84.22%	82.50%	84.22%	84.22%	84.22%	84.22%	84.22%	82.50%	84.22%	84.22%	84.22%	84.22%	84.22%	82.50%	84.22%
AVG100	14.91%	16.65%	54.67%	54.67%	52.58%	53.04%	54.01%	54.96%	54.67%	54.67%	52.58%	53.04%	54.01%	54.96%	54.67%	54.67%	52.58%	53.04%

T401 Query 2 entropy

T401 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	83.33%	100%	83.33%	100%	100%	83.33%	83.33%	83.33%	83.33%	100%	100%	83.33%	100%	100%	100%	100%	100%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	75%	81.82%	81.82%	81.82%	81.82%	81.82%	75%	81.82%	81.82%	81.82%	81.82%	81.82%	90%	81.82%
30%	87.50%	87.50%	87.50%	73.68%	82.35%	87.50%	87.50%	87.50%	87.50%	73.68%	82.35%	87.50%	87.50%	73.68%	73.68%	73.68%	82.35%	87.50%
40%	81.82%	78.26%	85.71%	72%	72%	81.82%	81.82%	85.71%	85.71%	72%	72%	81.82%	81.82%	72%	72%	72%	72%	81.82%
50%	82.14%	79.31%	82.14%	71.88%	71.88%	71.88%	71.88%	82.14%	82.14%	71.88%	71.88%	71.88%	71.88%	71.88%	71.88%	71.88%	71.88%	71.88%
60%	81.82%	79.41%	84.38%	71.05%	72.97%	72.97%	72.97%	84.38%	84.38%	71.05%	72.97%	72.97%	72.97%	71.05%	71.05%	71.05%	72.97%	72.97%
70%	71.11%	72.73%	72.73%	66.67%	68.09%	58.18%	65.31%	72.73%	72.73%	66.67%	68.09%	58.18%	64%	66.67%	66.67%	66.67%	68.09%	58.18%
80%	16.82%	63.16%	54.55%	42.86%	36.74%	39.56%	40.45%	54.55%	54.55%	42.86%	36.74%	39.56%	40.45%	42.86%	42.86%	42.86%	36.74%	39.56%
90%	1.89%	8.63%	6.07%	3.97%	3.69%	1.87%	3.65%	5.41%	5.40%	3.40%	3.59%	1.56%	3.95%	3.40%	3.41%	3.40%	3.59%	1.56%
100%	1.01%	0.86%	0.86%	0.79%	0.75%	0.59%	0.66%	0.70%	0.70%	0.68%	0.64%	0.50%	0.61%	0.67%	0.67%	0.66%	0.64%	0.49%
Avg30	84.22%	89.77%	84.22%	85.17%	85.78%	84.22%	84.22%	84.22%	84.22%	85.17%	85.78%	84.22%	89.77%	85.17%	85.17%	85.17%	90.78%	84.22%
AVG100	58.93%	65.17%	63.91%	58.47%	58.35%	57.95%	58.94%	63.83%	63.83%	58.40%	58.32%	57.91%	60.50%	58.40%	58.40%	58.40%	59.82%	57.91%

T401 Query 3 entropy

T401 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	93.33%	87.50%	87.50%	93.33%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%
40%	94.74%	94.74%	94.74%	90%	81.82%	85.71%	94.74%	94.74%	94.74%	90%	81.82%	85.71%	85.71%	90%	90%	90%	81.82%	85.71%
50%	82.14%	82.14%	82.14%	92%	79.31%	82.14%	82.14%	82.14%	82.14%	92%	79.31%	82.14%	82.14%	92%	92%	92%	79.31%	82.14%
60%	84.38%	84.38%	84.38%	84.38%	81.82%	81.82%	84.38%	84.38%	84.38%	84.38%	81.82%	81.82%	81.82%	84.38%	84.38%	84.38%	81.82%	81.82%
70%	86.49%	86.49%	86.49%	84.21%	80%	84.21%	86.49%	86.49%	86.49%	84.21%	80%	84.21%	84.21%	84.21%	84.21%	84.21%	80%	84.21%
80%	69.23%	70.59%	70.59%	70.59%	78.26%	73.47%	69.23%	70.59%	70.59%	70.59%	78.26%	73.47%	73.47%	70.59%	70.59%	70.59%	78.26%	73.47%
90%	0.80%	0.98%	0.98%	0.96%	1.13%	0.99%	0.80%	0.98%	0.98%	0.96%	1.13%	0.99%	0.79%	0.95%	0.95%	0.96%	1.13%	0.99%
100%	0.51%	0.57%	0.57%	0.56%	0.62%	0.19%	0.51%	0.57%	0.57%	0.56%	0.62%	0.19%	0.51%	0.56%	0.56%	0.55%	0.62%	0.18%
Avg30	94.44%	94.44%	94.44%	94.44%	92.50%	92.50%	94.44%	94.44%	94.44%	94.44%	92.50%	92.50%	92.50%	94.44%	94.44%	94.44%	92.50%	92.50%
AVG100	70.16%	70.32%	70.32%	70.60%	68.05%	68.60%	70.16%	70.32%	70.32%	70.60%	68.05%	68.60%	68.62%	70.60%	70.60%	70.60%	68.05%	68.60%

T401 Query 4 entropy

T401 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	83.33%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	81.82%	90%	90%	81.82%	90%	90%	90%	90%	90%	81.82%	81.82%	81.82%	81.82%	90%	90%	90%	90%	81.82%
30%	87.50%	82.35%	82.35%	87.50%	87.50%	87.50%	82.35%	82.35%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	85.71%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
50%	88.46%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%
60%	87.10%	84.38%	84.38%	87.10%	87.10%	84.38%	84.38%	84.38%	84.38%	87.10%	87.10%	84.38%	84.38%	87.10%	87.10%	87.10%	87.10%	84.38%
70%	80%	78.05%	78.05%	84.21%	84.21%	84.21%	82.05%	78.05%	78.05%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	23.53%	67.93%	67.93%	69.23%	70.59%	64.29%	65.46%	67.93%	67.93%	69.23%	70.59%	64.29%	69.23%	67.93%	67.93%	69.23%	70.59%	64.29%
90%	1.65%	5.69%	5.69%	7.39%	4.64%	0.83%	5.24%	5.69%	5.69%	7.32%	4.64%	0.77%	6.14%	6.61%	6.61%	7.32%	4.64%	0.77%
100%	0.74%	0.73%	0.73%	0.62%	0.52%	0.45%	0.60%	0.64%	0.64%	0.59%	0.46%	0.40%	0.53%	0.56%	0.56%	0.57%	0.41%	0.34%
Avg30	84.22%	90.78%	90.78%	89.77%	92.50%	92.50%	90.78%	90.78%	90.78%	89.77%	89.77%	89.77%	89.77%	92.50%	92.50%	92.50%	92.50%	89.77%
AVG100	61.99%	67.61%	67.61%	68.49%	69.16%	67.87%	67.71%	67.60%	67.60%	68.48%	68.33%	67.04%	68.08%	69.09%	69.09%	69.29%	69.15%	67.03%

T401 Query 5 entropy

T401 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%
40%	85.71%	72%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%
50%	88.46%	74.19%	71.88%	71.88%	79.31%	88.46%	82.14%	74.19%	71.88%	71.88%	79.31%	88.46%	82.14%	74.19%	71.88%	71.88%	79.31%	88.46%
60%	87.10%	77.14%	72.97%	72.97%	79.41%	87.10%	81.82%	75%	72.97%	72.97%	79.41%	87.10%	81.82%	75%	72.97%	72.97%	79.41%	87.10%
70%	71.11%	78.05%	68.09%	68.09%	60.38%	88.89%	68.09%	69.57%	68.09%	68.09%	60.38%	88.89%	68.09%	69.57%	68.09%	68.09%	60.38%	88.89%
80%	12.63%	16.22%	10.81%	10.81%	26.28%	40%	9.86%	10.75%	10.78%	10.78%	26.28%	40%	9.86%	10.75%	10.78%	10.78%	26.28%	40%
90%	0.97%	1.25%	2.49%	2.49%	2.35%	1.03%	1.98%	2.48%	2.43%	2.43%	2.35%	1.03%	1.98%	2.48%	2.43%	2.43%	2.35%	1.03%
100%	0.62%	0.73%	0.69%	0.69%	0.74%	0.63%	0.59%	0.68%	0.67%	0.67%	0.73%	0.55%	0.59%	0.68%	0.67%	0.67%	0.73%	0.55%
Avg30	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%
AVG100	59.41%	56.71%	56.01%	56.01%	57.32%	64.45%	57.77%	56.59%	56.00%	56.00%	57.32%	64.44%	57.77%	56.59%	56.00%	56.00%	57.32%	64.44%

T401 Query 6 entropy

T401 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	64.29%	60%	60%	64.29%	62.07%	85.71%	60%	60%	60%	64.29%	62.07%	85.71%	85.71%	64.29%	64.29%	64.29%	62.07%	85.71%
50%	63.89%	63.89%	63.89%	62.16%	62.16%	65.71%	63.89%	63.89%	63.89%	62.16%	62.16%	65.71%	51.11%	62.16%	62.16%	62.16%	62.16%	65.71%
60%	61.36%	55.10%	55.10%	55.10%	62.79%	67.50%	64.29%	55.10%	55.10%	55.10%	62.79%	67.50%	52.94%	55.10%	55.10%	55.10%	62.79%	67.50%
70%	59.26%	51.61%	51.61%	53.33%	66.67%	71.11%	64%	51.61%	51.61%	53.33%	66.67%	71.11%	55.17%	53.33%	53.33%	53.33%	66.67%	71.11%
80%	19.36%	16%	16%	28.80%	31.30%	38.71%	22.36%	14.40%	14.40%	27.07%	29.27%	38.71%	26.09%	27.69%	27.69%	27.07%	29.27%	38.71%
90%	1.17%	1.25%	1.25%	7.00%	5.53%	1.56%	1.17%	1.24%	1.24%	6.76%	5.47%	1.54%	6.30%	6.85%	6.85%	6.43%	5.45%	1.54%
100%	0.66%	0.71%	0.71%	0.82%	0.84%	0.77%	0.66%	0.70%	0.70%	0.82%	0.84%	0.77%	0.77%	0.77%	0.77%	0.75%	0.82%	0.75%
Avg30	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%
AVG100	55.33%	53.19%	53.19%	55.48%	57.47%	61.44%	55.97%	53.03%	53.03%	55.29%	57.26%	61.44%	56.14%	55.35%	55.35%	55.25%	57.26%	61.44%

T401 Query 7 entropy

T401 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	83.33%	83.33%	83.33%	100%	100%	100%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	36%	39.13%	39.13%	90%	90%	90%	36%	39.13%	39.13%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	18.67%	16.67%	16.67%	93.33%	93.33%	93.33%	18.67%	16.67%	16.67%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	15.13%	18%	18%	94.74%	90%	90%	15.13%	18%	18%	94.74%	90%	90%	90%	94.74%	94.74%	94.74%	90%	90%
50%	15.33%	15.03%	15.03%	88.46%	79.31%	82.14%	15.33%	15.03%	15.03%	88.46%	79.31%	82.14%	88.46%	88.46%	88.46%	88.46%	79.31%	82.14%
60%	12.22%	11.44%	11.44%	84.38%	72.97%	81.82%	12.22%	11.44%	11.44%	84.38%	72.97%	81.82%	84.38%	90%	90%	84.38%	72.97%	81.82%
70%	9.91%	8.16%	8.16%	66.67%	49.23%	55.17%	9.91%	8.16%	8.16%	66.67%	49.23%	55.17%	64%	69.57%	69.57%	66.67%	49.23%	55.17%
80%	2.69%	3.15%	3.15%	47.37%	33.65%	37.11%	2.69%	3.15%	3.15%	47.37%	33.65%	37.11%	43.37%	48%	48%	47.37%	33.65%	37.11%
90%	1.10%	1.32%	1.32%	1.00%	1.12%	1.20%	1.10%	1.32%	1.32%	1.00%	1.12%	1.20%	0.73%	1.00%	1.00%	0.91%	1.12%	1.14%
100%	0.57%	0.65%	0.65%	0.62%	0.65%	0.20%	0.57%	0.65%	0.65%	0.62%	0.65%	0.20%	0.50%	0.62%	0.62%	0.55%	0.65%	0.19%
Avg30	51.56%	51.93%	51.93%	88.89%	88.89%	88.89%	51.56%	51.93%	51.93%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%
AVG100	21.16%	21.35%	21.35%	64.99%	59.36%	61.43%	21.16%	21.35%	21.35%	64.99%	59.36%	61.43%	63.81%	65.91%	65.91%	64.97%	59.36%	61.42%

T401 Query 8 entropy

T401 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	100%	100%	100%	100%	100%	90%	90%	100%	100%	100%	100%	90%	90%	100%	100%	100%	100%	90%
30%	93.33%	93.33%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%
40%	75%	78.26%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	54.76%	53.49%	88.46%	88.46%	85.19%	88.46%	88.46%	88.46%	88.46%	88.46%	85.19%	88.46%	88.46%	88.46%	88.46%	88.46%	85.19%	88.46%
60%	48.21%	46.55%	87.10%	87.10%	87.10%	90%	90%	87.10%	87.10%	87.10%	87.10%	90%	90%	87.10%	87.10%	87.10%	87.10%	90%
70%	36.78%	44.44%	78.05%	78.05%	78.05%	78.05%	78.05%	80%	78.05%	78.05%	78.05%	78.05%	78.05%	80%	78.05%	78.05%	78.05%	78.05%
80%	24%	27.91%	42.35%	42.35%	58.07%	40%	42.86%	42.35%	42.35%	42.35%	58.07%	40%	42.86%	42.35%	42.35%	42.35%	58.07%	40%
90%	2.74%	1.76%	3.59%	3.59%	2.76%	0.81%	2.61%	3.35%	3.59%	3.59%	2.76%	0.81%	2.61%	3.35%	3.59%	3.59%	2.76%	0.81%
100%	0.63%	0.72%	0.75%	0.75%	0.82%	0.44%	0.64%	0.71%	0.71%	0.71%	0.80%	0.41%	0.64%	0.71%	0.71%	0.71%	0.80%	0.41%
Avg30	97.78%	97.78%	95.83%	95.83%	97.78%	94.44%	94.44%	95.83%	95.83%	95.83%	97.78%	94.44%	94.44%	95.83%	95.83%	95.83%	97.78%	94.44%
AVG100	53.55%	54.65%	67.35%	67.35%	69.10%	66.68%	67.17%	67.52%	67.35%	67.35%	69.10%	66.68%	67.17%	67.52%	67.35%	67.35%	69.10%	66.68%

T401 Query 9 entropy

T401 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	75%	81.82%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	88.46%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%
60%	84.38%	87.10%	87.10%	87.10%	87.10%	84.38%	87.10%	87.10%	87.10%	87.10%	87.10%	84.38%	87.10%	87.10%	87.10%	87.10%	87.10%	84.38%
70%	78.05%	80%	80%	84.21%	84.21%	84.21%	82.05%	80%	80%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	25.90%	35.29%	35.29%	61.02%	75%	72%	46.15%	35.29%	35.29%	61.02%	75%	72%	70.59%	60%	60%	61.02%	75%	72%
90%	3.39%	2.03%	2.03%	1.99%	0.63%	0.47%	1.62%	2.02%	2.02%	1.98%	0.58%	0.44%	1.31%	1.71%	1.71%	1.96%	0.48%	0.32%
100%	0.97%	0.85%	0.85%	0.79%	0.49%	0.45%	0.73%	0.78%	0.78%	0.76%	0.46%	0.42%	0.66%	0.74%	0.74%	0.70%	0.35%	0.31%
Avg30	85.78%	89.77%	89.77%	92.50%	92.50%	92.50%	89.77%	89.77%	89.77%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%	92.50%
AVG100	62.09%	64.55%	64.55%	68.35%	69.58%	68.99%	66.11%	64.54%	64.54%	68.35%	69.57%	68.98%	69.23%	68.21%	68.21%	68.34%	69.55%	68.96%

T401 Query 10 entropy

T401 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	83.33%	83.33%	100%	100%	100%	100%	83.33%	83.33%	100%	100%	100%	100%	83.33%	83.33%
20%	90%	90%	81.82%	81.82%	90%	90%	81.82%	81.82%	81.82%	81.82%	90%	90%	81.82%	81.82%	81.82%	81.82%	90%	90%
30%	82.35%	87.50%	82.35%	82.35%	87.50%	87.50%	82.35%	82.35%	82.35%	82.35%	87.50%	87.50%	82.35%	82.35%	82.35%	82.35%	87.50%	87.50%
40%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%
50%	88.46%	88.46%	85.19%	85.19%	85.19%	85.19%	88.46%	88.46%	85.19%	85.19%	85.19%	85.19%	88.46%	88.46%	85.19%	85.19%	85.19%	85.19%
60%	87.10%	84.38%	87.10%	87.10%	81.82%	84.38%	87.10%	87.10%	87.10%	87.10%	81.82%	84.38%	87.10%	87.10%	87.10%	87.10%	81.82%	84.38%
70%	69.57%	69.57%	69.57%	69.57%	71.11%	74.42%	69.57%	69.57%	69.57%	69.57%	71.11%	74.42%	69.57%	69.57%	69.57%	69.57%	71.11%	74.42%
80%	34.29%	36.74%	31.03%	31.03%	63.16%	62.07%	30%	32.43%	30.77%	30.77%	63.16%	62.07%	30%	32.43%	30.77%	30.77%	63.16%	62.07%
90%	0.71%	0.83%	1.58%	1.58%	4.08%	0.51%	1.34%	1.59%	1.55%	1.55%	4.08%	0.51%	1.34%	1.59%	1.55%	1.55%	4.08%	0.51%
100%	0.45%	0.50%	0.50%	0.50%	0.64%	0.36%	0.43%	0.49%	0.48%	0.48%	0.64%	0.34%	0.43%	0.49%	0.48%	0.48%	0.64%	0.34%
Avg30	90.78%	92.50%	88.06%	88.06%	86.94%	86.94%	88.06%	88.06%	88.06%	88.06%	86.94%	86.94%	88.06%	88.06%	88.06%	88.06%	86.94%	86.94%
AVG100	63.86%	64.37%	62.48%	62.48%	64.86%	65.35%	62.68%	62.95%	62.45%	62.45%	64.86%	65.34%	62.68%	62.95%	62.45%	62.45%	64.86%	65.34%

T401 Query 1 idf

T401 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	38.46%	27.78%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	34.62%	34.62%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	28.57%	38.89%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%
40%	18.18%	24.32%	85.71%	85.71%	72%	85.71%	85.71%	85.71%	85.71%	85.71%	72%	85.71%	85.71%	85.71%	85.71%	85.71%	72%	85.71%
50%	16.79%	21.70%	82.14%	82.14%	67.65%	76.67%	82.14%	82.14%	82.14%	82.14%	67.65%	76.67%	82.14%	82.14%	82.14%	82.14%	67.65%	76.67%
60%	11.49%	13.64%	72.97%	72.97%	64.29%	75%	77.14%	72.97%	72.97%	72.97%	64.29%	75%	77.14%	72.97%	72.97%	72.97%	64.29%	75%
70%	3.39%	8.42%	42.11%	42.11%	25.81%	25.81%	30.77%	42.11%	42.11%	42.11%	25.81%	25.81%	30.77%	42.11%	42.11%	42.11%	25.81%	25.81%
80%	0.73%	0.84%	11.36%	11.36%	8.85%	7.69%	8.47%	12.33%	11.36%	11.36%	8.85%	7.69%	8.47%	12.33%	11.36%	11.36%	8.85%	7.69%
90%	0.67%	0.68%	1.06%	1.06%	1.24%	1.21%	0.83%	1.02%	1.02%	1.02%	1.24%	1.19%	0.83%	1.02%	1.02%	1.02%	1.24%	1.19%
100%	0.32%	0.45%	0.49%	0.49%	0.15%	0.14%	0.45%	0.48%	0.46%	0.46%	0.15%	0.13%	0.45%	0.48%	0.46%	0.46%	0.15%	0.13%
Avg30	33.88%	33.76%	84.22%	84.22%	82.50%	84.22%	84.22%	84.22%	84.22%	84.22%	82.50%	84.22%	84.22%	84.22%	84.22%	84.22%	82.50%	84.22%
AVG100	15.32%	17.13%	54.85%	54.85%	48.75%	52.49%	53.82%	54.94%	54.84%	54.84%	48.75%	52.48%	53.82%	54.94%	54.84%	54.84%	48.75%	52.48%

T401 Query 2 idf

T401 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	83.33%	83.33%	83.33%	83.33%	100%	83.33%	83.33%	83.33%	83.33%	83.33%	100%	83.33%	83.33%	83.33%	83.33%	83.33%	100%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	82.35%	87.50%
40%	90%	78.26%	85.71%	78.26%	85.71%	81.82%	81.82%	78.26%	85.71%	78.26%	85.71%	81.82%	81.82%	78.26%	78.26%	78.26%	85.71%	81.82%
50%	82.14%	79.31%	76.67%	76.67%	82.14%	71.88%	71.88%	76.67%	76.67%	76.67%	82.14%	71.88%	71.88%	76.67%	76.67%	76.67%	82.14%	71.88%
60%	81.82%	79.41%	79.41%	79.41%	77.14%	72.97%	75%	79.41%	79.41%	79.41%	77.14%	72.97%	75%	79.41%	79.41%	79.41%	77.14%	72.97%
70%	72.73%	72.73%	71.11%	71.11%	68.09%	58.18%	65.31%	71.11%	71.11%	71.11%	68.09%	58.18%	62.75%	71.11%	71.11%	71.11%	68.09%	58.18%
80%	16.74%	63.16%	55.39%	55.39%	42.86%	41.38%	42.35%	55.39%	55.39%	55.39%	42.86%	41.38%	42.35%	55.39%	55.39%	55.39%	42.86%	41.38%
90%	1.87%	9.09%	5.28%	4.90%	3.88%	1.91%	3.66%	4.70%	4.70%	4.38%	3.81%	1.61%	4.05%	4.38%	4.38%	4.38%	3.81%	1.61%
100%	1.01%	0.86%	0.85%	0.79%	0.73%	0.59%	0.66%	0.70%	0.70%	0.68%	0.62%	0.49%	0.61%	0.67%	0.67%	0.66%	0.62%	0.49%
Avg30	84.22%	84.22%	84.22%	84.22%	88.06%	84.22%	84.22%	84.22%	84.22%	84.22%	88.06%	84.22%	84.22%	84.22%	84.22%	84.22%	88.06%	84.22%
AVG100	59.90%	63.55%	62.71%	61.92%	62.47%	58.14%	59.33%	61.89%	62.63%	61.85%	62.45%	58.10%	59.11%	61.85%	61.85%	61.85%	62.45%	58.10%

T401 Query 3 idf

T401 Q3	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	93.33%	93.33%	87.50%	93.33%	93.33%	93.33%	93.33%	93.33%	87.50%	87.50%	93.33%	93.33%	93.33%	93.33%	87.50%
40%	94.74%	94.74%	94.74%	90%	85.71%	85.71%	94.74%	94.74%	94.74%	90%	85.71%	85.71%	85.71%	90%	90%	90%	85.71%	85.71%
50%	82.14%	82.14%	82.14%	92%	88.46%	88.46%	82.14%	82.14%	82.14%	92%	88.46%	88.46%	88.46%	92%	92%	92%	88.46%	88.46%
60%	84.38%	84.38%	84.38%	84.38%	81.82%	81.82%	84.38%	84.38%	84.38%	84.38%	81.82%	81.82%	81.82%	84.38%	84.38%	84.38%	81.82%	81.82%
70%	86.49%	86.49%	86.49%	84.21%	82.05%	82.05%	86.49%	86.49%	86.49%	84.21%	82.05%	82.05%	84.21%	84.21%	84.21%	84.21%	82.05%	82.05%
80%	69.23%	69.23%	69.23%	69.23%	75%	73.47%	69.23%	69.23%	69.23%	69.23%	75%	73.47%	70.59%	69.23%	69.23%	69.23%	75%	73.47%
90%	0.77%	0.96%	0.96%	0.93%	1.08%	0.97%	0.77%	0.96%	0.96%	0.93%	1.08%	0.97%	0.77%	0.93%	0.93%	0.93%	1.08%	0.97%
100%	0.50%	0.55%	0.55%	0.54%	0.62%	0.19%	0.50%	0.55%	0.55%	0.54%	0.62%	0.19%	0.50%	0.54%	0.54%	0.54%	0.62%	0.19%
Avg30	94.44%	94.44%	94.44%	94.44%	94.44%	92.50%	94.44%	94.44%	94.44%	94.44%	94.44%	92.50%	92.50%	94.44%	94.44%	94.44%	94.44%	92.50%
AVG100	70.16%	70.18%	70.18%	70.46%	69.81%	69.02%	70.16%	70.18%	70.18%	70.46%	69.81%	69.02%	68.96%	70.46%	70.46%	70.46%	69.81%	69.02%

T401 Query 4 idf

T401 Q4	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	83.33%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	81.82%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	87.50%	82.35%	82.35%	87.50%	87.50%	87.50%	87.50%	82.35%	82.35%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%	87.50%
40%	85.71%	81.82%	81.82%	81.82%	81.82%	81.82%	85.71%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
50%	88.46%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%	85.19%
60%	87.10%	84.38%	84.38%	84.38%	87.10%	87.10%	87.10%	84.38%	84.38%	84.38%	87.10%	87.10%	87.10%	84.38%	84.38%	84.38%	87.10%	87.10%
70%	80%	78.05%	78.05%	84.21%	84.21%	84.21%	84.21%	78.05%	78.05%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	24.83%	67.93%	67.93%	69.23%	70.59%	65.46%	69.23%	67.93%	67.93%	69.23%	70.59%	65.46%	69.23%	69.23%	69.23%	69.23%	70.59%	65.46%
90%	1.74%	5.66%	5.66%	7.35%	4.61%	0.87%	4.90%	5.66%	5.66%	7.28%	4.61%	0.81%	6.16%	6.43%	6.43%	7.28%	4.61%	0.81%
100%	0.74%	0.72%	0.72%	0.62%	0.51%	0.45%	0.57%	0.64%	0.64%	0.59%	0.46%	0.40%	0.53%	0.57%	0.57%	0.57%	0.41%	0.34%
Avg30	84.22%	90.78%	90.78%	89.77%	89.77%	89.77%	92.50%	90.78%	90.78%	89.77%	89.77%	89.77%	89.77%	89.77%	89.77%	89.77%	89.77%	89.77%
AVG100	62.12%	67.61%	67.61%	68.21%	68.33%	67.44%	69.44%	67.60%	67.60%	68.20%	68.33%	67.43%	68.35%	68.11%	68.11%	68.20%	68.32%	67.42%

T401 Query 5 idf

T401 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
30%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%	82.35%	82.35%	82.35%	82.35%	77.78%	87.50%
40%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%
50%	88.46%	76.67%	82.14%	82.14%	79.31%	88.46%	82.14%	85.19%	82.14%	82.14%	79.31%	88.46%	82.14%	85.19%	82.14%	82.14%	79.31%	88.46%
60%	87.10%	79.41%	75%	75%	79.41%	87.10%	84.38%	75%	75%	75%	79.41%	87.10%	84.38%	75%	75%	75%	79.41%	87.10%
70%	72.73%	76.19%	68.09%	68.09%	62.75%	88.89%	68.09%	71.11%	68.09%	68.09%	62.75%	88.89%	68.09%	71.11%	68.09%	68.09%	62.75%	88.89%
80%	16.82%	12.12%	10.91%	10.91%	26.09%	38.30%	9.89%	10.88%	10.91%	10.91%	26.09%	38.30%	9.89%	10.88%	10.91%	10.91%	26.09%	38.30%
90%	1.04%	1.22%	2.43%	2.43%	2.46%	1.04%	1.97%	2.44%	2.37%	2.37%	2.46%	1.04%	1.97%	2.44%	2.37%	2.37%	2.46%	1.04%
100%	0.61%	0.69%	0.68%	0.68%	0.72%	0.63%	0.58%	0.67%	0.66%	0.66%	0.72%	0.55%	0.58%	0.67%	0.66%	0.66%	0.72%	0.55%
Avg30	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%	82.50%	82.50%	82.50%	82.50%	80.98%	84.22%
AVG100	60.00%	57.95%	57.25%	57.25%	57.55%	64.28%	58.03%	57.85%	57.24%	57.24%	57.55%	64.27%	58.03%	57.85%	57.24%	57.24%	57.55%	64.27%

T401 Query 6 idf

T401 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	64.29%	60%	60%	64.29%	62.07%	85.71%	62.07%	60%	60%	64.29%	62.07%	85.71%	85.71%	64.29%	64.29%	64.29%	62.07%	85.71%
50%	63.89%	63.89%	63.89%	62.16%	62.16%	65.71%	62.16%	63.89%	63.89%	62.16%	62.16%	65.71%	54.76%	62.16%	62.16%	62.16%	62.16%	65.71%
60%	62.79%	58.70%	58.70%	61.36%	62.79%	67.50%	64.29%	58.70%	58.70%	61.36%	62.79%	67.50%	52.94%	61.36%	61.36%	61.36%	62.79%	67.50%
70%	61.54%	58.18%	58.18%	57.14%	66.67%	71.11%	64%	58.18%	58.18%	57.14%	66.67%	71.11%	57.14%	56.14%	56.14%	57.14%	66.67%	71.11%
80%	19.57%	16.82%	16.82%	29.03%	31.03%	38.30%	23.84%	15.06%	15.06%	27.27%	29.03%	38.30%	29.75%	27.91%	27.91%	27.27%	29.03%	38.30%
90%	1.18%	1.27%	1.27%	7.00%	5.58%	1.56%	1.22%	1.25%	1.25%	6.76%	5.51%	1.54%	6.25%	6.86%	6.86%	6.50%	5.50%	1.54%
100%	0.68%	0.72%	0.72%	0.83%	0.85%	0.78%	0.69%	0.72%	0.72%	0.83%	0.84%	0.77%	0.78%	0.78%	0.78%	0.76%	0.82%	0.76%
Avg30	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%	94.44%
AVG100	55.73%	54.29%	54.29%	56.52%	57.45%	61.40%	56.16%	54.11%	54.11%	56.31%	57.24%	61.40%	57.07%	56.28%	56.28%	56.28%	57.24%	61.40%

T401 Query 7 idf

T401 Q7	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	83.33%	83.33%	83.33%	100%	100%	100%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%	83.33%
20%	36%	37.50%	37.50%	90%	90%	90%	36%	37.50%	37.50%	90%	90%	90%	90%	90%	90%	90%	90%	90%
30%	17.72%	16.47%	16.47%	93.33%	93.33%	93.33%	17.72%	16.47%	16.47%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%	93.33%
40%	15.25%	15.79%	15.79%	90%	90%	85.71%	15.25%	15.79%	15.79%	90%	90%	85.71%	85.71%	90%	90%	90%	90%	85.71%
50%	14.94%	15.23%	15.23%	92%	79.31%	79.31%	14.94%	15.23%	15.23%	92%	79.31%	79.31%	79.31%	92%	92%	92%	79.31%	79.31%
60%	13.04%	11.64%	11.64%	84.38%	72.97%	72.97%	13.04%	11.64%	11.64%	84.38%	72.97%	72.97%	79.41%	90%	90%	84.38%	72.97%	72.97%
70%	9.85%	8.31%	8.31%	66.67%	49.23%	53.33%	9.85%	8.31%	8.31%	66.67%	49.23%	53.33%	53.33%	69.57%	69.57%	66.67%	49.23%	53.33%
80%	2.83%	2.96%	2.96%	46.75%	32.73%	34.95%	2.83%	2.96%	2.96%	46.75%	32.73%	34.95%	36.74%	46.75%	46.75%	46.75%	32.73%	34.95%
90%	1.10%	1.26%	1.26%	0.98%	1.08%	1.18%	1.10%	1.26%	1.26%	0.98%	1.08%	1.18%	0.75%	0.97%	0.97%	0.90%	1.08%	1.13%
100%	0.55%	0.62%	0.62%	0.61%	0.64%	0.23%	0.55%	0.62%	0.62%	0.61%	0.64%	0.23%	0.50%	0.61%	0.61%	0.54%	0.64%	0.22%
Avg30	51.24%	51.32%	51.32%	88.89%	88.89%	88.89%	51.24%	51.32%	51.32%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%	88.89%
AVG100	21.13%	20.98%	20.98%	64.80%	59.26%	59.44%	21.13%	20.98%	20.98%	64.80%	59.26%	59.44%	60.24%	65.66%	65.66%	64.79%	59.26%	59.43%

T401 Query 8 idf

T401 Q8	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	100%	100%	100%	100%	100%	90%	90%	100%	100%	100%	100%	90%	90%	100%	100%	100%	100%	90%
30%	93.33%	93.33%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%	93.33%	87.50%	87.50%	87.50%	93.33%	93.33%
40%	75%	78.26%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
50%	54.76%	53.49%	88.46%	88.46%	85.19%	88.46%	88.46%	88.46%	88.46%	88.46%	85.19%	88.46%	88.46%	88.46%	88.46%	88.46%	85.19%	88.46%
60%	47.37%	46.55%	87.10%	87.10%	87.10%	90%	90%	87.10%	87.10%	87.10%	87.10%	90%	90%	87.10%	87.10%	87.10%	87.10%	90%
70%	36.36%	44.44%	78.05%	78.05%	78.05%	78.05%	78.05%	80%	78.05%	78.05%	78.05%	78.05%	78.05%	80%	78.05%	78.05%	78.05%	78.05%
80%	24%	27.69%	42.86%	42.86%	58.07%	40%	41.86%	42.86%	42.86%	42.86%	58.07%	40%	41.86%	42.86%	42.86%	42.86%	58.07%	40%
90%	2.82%	1.95%	3.52%	3.52%	2.72%	0.84%	2.56%	3.30%	3.52%	3.52%	2.72%	0.84%	2.56%	3.30%	3.52%	3.52%	2.72%	0.84%
100%	0.63%	0.71%	0.75%	0.75%	0.81%	0.43%	0.64%	0.71%	0.71%	0.71%	0.79%	0.41%	0.64%	0.71%	0.71%	0.71%	0.79%	0.41%
Avg30	97.78%	97.78%	95.83%	95.83%	97.78%	94.44%	94.44%	95.83%	95.83%	95.83%	97.78%	94.44%	94.44%	95.83%	95.83%	95.83%	97.78%	94.44%
AVG100	53.43%	54.64%	67.39%	67.39%	69.10%	66.68%	67.06%	67.56%	67.39%	67.39%	69.10%	66.68%	67.06%	67.56%	67.39%	67.39%	69.10%	66.68%

T401 Query 9 idf

T401 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	75%	81.82%	81.82%	90%	100%	90%	81.82%	81.82%	81.82%	90%	100%	90%	100%	90%	90%	90%	100%	90%
30%	82.35%	87.50%	87.50%	87.50%	77.78%	87.50%	87.50%	87.50%	87.50%	87.50%	77.78%	87.50%	77.78%	87.50%	87.50%	87.50%	77.78%	87.50%
40%	85.71%	85.71%	85.71%	85.71%	78.26%	85.71%	85.71%	85.71%	85.71%	85.71%	78.26%	85.71%	78.26%	85.71%	85.71%	85.71%	78.26%	85.71%
50%	85.19%	88.46%	88.46%	85.19%	82.14%	85.19%	88.46%	88.46%	88.46%	85.19%	82.14%	85.19%	82.14%	85.19%	85.19%	85.19%	82.14%	85.19%
60%	84.38%	87.10%	87.10%	87.10%	81.82%	87.10%	87.10%	87.10%	87.10%	87.10%	81.82%	87.10%	81.82%	87.10%	87.10%	87.10%	81.82%	87.10%
70%	78.05%	80%	80%	84.21%	84.21%	84.21%	82.05%	80%	80%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%	84.21%
80%	26.28%	35.64%	35.64%	61.02%	75%	72%	46.15%	35.64%	35.64%	61.02%	75%	72%	75%	61.02%	61.02%	61.02%	75%	72%
90%	3.40%	2.03%	2.03%	1.99%	0.66%	0.48%	1.62%	2.02%	2.02%	1.98%	0.60%	0.44%	1.26%	1.71%	1.71%	1.96%	0.50%	0.33%
100%	0.97%	0.85%	0.85%	0.79%	0.50%	0.46%	0.74%	0.78%	0.78%	0.77%	0.47%	0.42%	0.65%	0.74%	0.74%	0.71%	0.37%	0.31%
Avg30	85.78%	89.77%	89.77%	92.50%	92.59%	92.50%	89.77%	89.77%	89.77%	92.50%	92.59%	92.50%	92.59%	92.50%	92.50%	92.50%	92.59%	92.50%
AVG100	62.13%	64.91%	64.91%	68.35%	68.04%	69.26%	66.11%	64.90%	64.90%	68.35%	68.03%	69.26%	68.11%	68.32%	68.32%	68.34%	68.01%	69.23%

T401 Query 10 idf

T401 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	83.33%	83.33%	83.33%	100%	100%	100%	83.33%	83.33%	83.33%	100%	100%	100%	83.33%	83.33%
20%	90%	90%	81.82%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	90%	90%
30%	82.35%	87.50%	82.35%	82.35%	87.50%	87.50%	87.50%	82.35%	82.35%	82.35%	87.50%	87.50%	87.50%	82.35%	82.35%	82.35%	87.50%	87.50%
40%	85.71%	85.71%	85.71%	85.71%	81.82%	85.71%	90%	85.71%	85.71%	85.71%	81.82%	85.71%	90%	85.71%	85.71%	85.71%	81.82%	85.71%
50%	88.46%	88.46%	88.46%	88.46%	85.19%	85.19%	92%	88.46%	88.46%	88.46%	85.19%	85.19%	92%	88.46%	88.46%	88.46%	85.19%	85.19%
60%	87.10%	84.38%	87.10%	87.10%	81.82%	84.38%	90%	87.10%	87.10%	87.10%	81.82%	84.38%	90%	87.10%	87.10%	87.10%	81.82%	84.38%
70%	69.57%	66.67%	69.57%	69.57%	71.11%	74.42%	82.05%	69.57%	69.57%	69.57%	71.11%	74.42%	82.05%	69.57%	69.57%	69.57%	71.11%	74.42%
80%	33.33%	32.14%	31.30%	31.30%	63.16%	62.07%	72%	31.86%	31.03%	31.03%	63.16%	62.07%	72%	31.86%	31.03%	31.03%	63.16%	62.07%
90%	0.76%	0.87%	1.55%	1.55%	4%	0.51%	2.86%	1.56%	1.51%	1.51%	4%	0.51%	2.86%	1.56%	1.51%	1.51%	4%	0.51%
100%	0.45%	0.49%	0.49%	0.49%	0.63%	0.36%	0.49%	0.48%	0.48%	0.48%	0.63%	0.34%	0.49%	0.48%	0.48%	0.48%	0.63%	0.34%
Avg30	90.78%	92.50%	88.06%	88.06%	86.94%	86.94%	86.94%	88.06%	88.06%	88.06%	86.94%	86.94%	86.94%	88.06%	88.06%	88.06%	86.94%	86.94%
AVG100	63.77%	63.62%	62.84%	62.84%	64.86%	65.35%	69.02%	62.89%	62.80%	62.80%	64.86%	65.34%	69.02%	62.89%	62.80%	62.80%	64.86%	65.34%

T403 Query 1

T403 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	90%	90%	90%	90%	75%	90%	90%	90%	90%	90%	75%	90%	90%	81.82%	90%	90%	75%	90%
20%	62.07%	66.67%	72%	66.67%	69.23%	66.67%	66.67%	75%	72%	66.67%	69.23%	66.67%	60%	81.82%	66.67%	66.67%	69.23%	66.67%
30%	65.85%	65.85%	58.70%	46.55%	42.86%	45%	56.25%	62.79%	58.70%	46.55%	42.86%	45%	45.76%	57.45%	47.37%	46.55%	42.86%	45%
40%	46.75%	54.55%	55.39%	40.45%	39.13%	33.33%	45%	60%	55.39%	40.45%	39.13%	33.33%	42.86%	45%	39.56%	40.45%	39.13%	33.33%
50%	40%	39.32%	40.71%	36.80%	37.40%	28.40%	37.10%	43.81%	40.71%	36.80%	37.40%	28.40%	37.10%	33.82%	37.71%	36.80%	37.40%	28.40%
60%	26.32%	27.23%	22.54%	24.12%	28.50%	26.70%	25.11%	27.09%	22.54%	24.02%	28.50%	26.70%	26.32%	27.78%	24.44%	24.02%	28.50%	26.70%
70%	18.39%	18.50%	21.84%	18.39%	16.24%	19.10%	21.62%	24.15%	21.84%	18.23%	16.20%	19.10%	16.50%	19.10%	18.34%	18.23%	16.20%	19.10%
80%	11.18%	12.59%	20.51%	13.93%	6.80%	15.60%	19.42%	21.53%	20.51%	13.75%	6.76%	15.57%	13.30%	16.37%	13.85%	13.75%	6.76%	15.57%
90%	9.84%	10.31%	17.30%	10.42%	5.28%	11.71%	14.96%	16.50%	17.30%	10.25%	5.24%	11.70%	7.00%	11.31%	10.24%	10.24%	5.23%	11.70%
100%	0%	0%	0%	0%	0%	0%	5.44%	0.74%	5.48%	2.45%	1.29%	1.71%	3.28%	0.71%	2.77%	2.36%	1.22%	1.66%
Avg30	72.64%	74.17%	73.57%	67.74%	62.36%	67.22%	70.97%	75.93%	73.57%	67.74%	62.36%	67.22%	65.25%	73.69%	68.01%	67.74%	62.36%	67.22%
AVG100	37.04%	38.50%	39.90%	34.73%	32.04%	33.65%	38.16%	42.16%	40.45%	34.92%	32.16%	33.82%	34.21%	37.52%	35.09%	34.91%	32.15%	33.81%

T403 Query 2

T403 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	75%	75%	81.82%	75%	81.82%	75%	81.82%	81.82%	81.82%	75%	81.82%	75%	75%	75%	75%	75%	81.82%	75%
20%	72%	75%	64.29%	60%	60%	62.07%	60%	66.67%	64.29%	60%	60%	62.07%	54.55%	69.23%	64.29%	60%	60%	62.07%
30%	50%	50%	57.45%	61.36%	56.25%	60%	55.10%	58.70%	57.45%	61.36%	56.25%	60%	60%	62.79%	62.79%	61.36%	56.25%	60%
40%	36%	33.96%	41.38%	54.55%	54.55%	52.94%	42.86%	41.38%	41.38%	54.55%	54.55%	52.94%	53.73%	53.73%	54.55%	54.55%	54.55%	52.94%
50%	27.88%	15.97%	28.75%	38.66%	36.80%	35.12%	27.38%	25.28%	28.75%	38.66%	36.80%	35.12%	39.32%	39.66%	38.98%	38.66%	36.80%	35.12%
60%	14.25%	13.32%	16.04%	33.74%	34.16%	30.90%	20.91%	14.67%	16.04%	33.74%	34.16%	30.90%	37.16%	27.09%	27.92%	33.74%	34.16%	30.90%
70%	9.80%	10.61%	8.27%	13.85%	15.57%	13.53%	11.35%	8.39%	8.27%	13.73%	15.46%	13.50%	15.06%	14.85%	14.04%	13.70%	15.46%	13.50%
80%	8.03%	5.39%	7.25%	8.42%	8.08%	8.74%	6.51%	7.26%	7.21%	8.31%	8.00%	8.64%	7.46%	8.65%	8.06%	8.29%	8.00%	8.64%
90%	2.82%	2.64%	3.67%	6.09%	2.89%	7.17%	3.18%	3.62%	3.52%	5.99%	2.82%	7.10%	5.13%	5.80%	5.84%	5.98%	2.81%	7.10%
100%	1.41%	1.41%	1.82%	1.01%	1.22%	2.73%	1.55%	2.03%	1.58%	0.95%	1.10%	2.52%	0.65%	1.87%	1.48%	0.66%	1.03%	2.51%
Avg30	65.67%	66.67%	67.85%	65.45%	66.02%	65.69%	65.64%	69.06%	67.85%	65.45%	66.02%	65.69%	63.18%	69.01%	67.36%	65.45%	66.02%	65.69%
AVG100	29.72%	28.33%	31.07%	35.27%	35.13%	34.82%	31.07%	30.98%	31.03%	35.23%	35.10%	34.78%	34.80%	35.87%	35.29%	35.19%	35.09%	34.78%

T403 Query 3

T403 Q3	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	75%	69.23%	64.29%	81.82%	81.82%	75%	69.23%	75%	69.23%	81.82%	81.82%	75%	81.82%	90%	81.82%	81.82%	81.82%	75%
20%	42.86%	38.30%	50%	81.82%	75%	81.82%	42.86%	62.07%	46.15%	81.82%	75%	81.82%	75%	78.26%	81.82%	81.82%	75%	81.82%
30%	29.03%	35.07%	51.92%	67.50%	58.70%	72.97%	47.37%	60%	50%	65.85%	58.70%	72.97%	65.85%	77.14%	69.23%	65.85%	58.70%	72.97%
40%	23.84%	34.29%	41.86%	33.65%	51.43%	36%	40.91%	53.73%	45%	31.30%	50%	35.64%	31.58%	35.64%	31.58%	31.30%	50%	35.64%
50%	9.18%	23.35%	36.51%	25.28%	29.30%	34.33%	36.80%	41.44%	38.98%	22.66%	27.88%	33.82%	21.50%	29.87%	22.66%	22.66%	27.88%	33.82%
60%	7.98%	20.30%	29.41%	26.07%	23.40%	27.36%	30.90%	29.10%	29.26%	23.11%	22.27%	26.57%	22.63%	26.57%	22.45%	23.11%	22.27%	26.57%
70%	7.79%	16.98%	24.24%	24.24%	14.65%	24.90%	25.50%	24.62%	23.53%	21.41%	13.97%	24.24%	23.88%	24.34%	20.19%	21.33%	13.97%	24.24%
80%	7.06%	11.99%	20.11%	21.92%	10.99%	20.17%	21.92%	16.15%	20.11%	18.91%	10.58%	19.52%	18.67%	19.42%	16.98%	18.81%	10.53%	19.52%
90%	5.11%	6.50%	4.52%	19.66%	10.49%	16.30%	4.52%	6.85%	8.01%	16.53%	9.93%	15.77%	15.65%	16.21%	14.94%	15.92%	9.90%	15.77%
100%	0%	0%	0%	0%	0%	0%	0.75%	4.01%	0.75%	5.91%	4.21%	3.70%	4.59%	3.04%	6.69%	4.80%	3.53%	3.43%
Avg30	48.96%	47.53%	55.40%	77.05%	71.84%	76.60%	53.15%	65.69%	55.13%	76.50%	71.84%	76.60%	74.22%	81.80%	77.62%	76.50%	71.84%	76.60%
AVG100	20.78%	25.60%	32.29%	38.20%	35.58%	38.89%	32.07%	37.30%	33.10%	36.93%	35.44%	38.91%	36.12%	40.05%	36.83%	36.74%	35.36%	38.88%

T403 Query 4

T403 Q4	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	69.23%	81.82%	75%	75%	60%	64.29%	69.23%	75%	75%	75%	60%	64.29%	75%	75%	75%	75%	60%	64.29%
20%	69.23%	72%	62.07%	56.25%	38.30%	39.13%	47.37%	69.23%	62.07%	56.25%	38.30%	39.13%	46.15%	66.67%	56.25%	56.25%	38.30%	39.13%
30%	40.30%	60%	48.21%	49.09%	44.26%	29.35%	39.71%	48.21%	48.21%	49.09%	44.26%	29.35%	38.57%	50.94%	50%	49.09%	44.26%	29.35%
40%	32.73%	60%	46.15%	39.13%	5.33%	24.83%	36.36%	50%	46.15%	39.13%	5.33%	24.83%	34.29%	47.37%	40.45%	39.13%	5.33%	24.83%
50%	25.70%	44.23%	34.33%	38.02%	5.53%	15.08%	31.29%	36.80%	34.07%	38.02%	5.53%	15.08%	29.68%	39.32%	38.98%	38.02%	5.53%	15.08%
60%	18.09%	39.57%	27.36%	26.32%	6.24%	13.06%	22.92%	31.98%	27.23%	26.19%	6.24%	13.06%	20.99%	27.64%	27.23%	26.19%	6.24%	13.06%
70%	14.25%	22.86%	16.71%	18.71%	3.82%	12.05%	19.69%	18.99%	16.54%	18.61%	3.82%	12.05%	18.88%	19.81%	19.57%	18.61%	3.82%	12.05%
80%	9.84%	17.14%	10.67%	9.49%	2.07%	11.61%	10.18%	12.15%	10.44%	9.37%	2.07%	11.61%	11.87%	13.11%	9.75%	9.35%	2.06%	11.61%
90%	7.11%	10.54%	6.45%	7.20%	2.07%	6.66%	8.16%	7.12%	6.22%	7.09%	2.05%	6.66%	8.38%	10.39%	7.66%	7.07%	2.05%	6.66%
100%	2.37%	2.60%	2.59%	2.61%	2.02%	2.05%	2.64%	2.60%	2.55%	2.57%	2.00%	2.04%	2.61%	2.60%	2.57%	2.55%	1.99%	2.03%
Avg30	59.59%	71.27%	61.76%	60.11%	47.52%	44.25%	52.10%	64.15%	61.76%	60.11%	47.52%	44.25%	53.24%	64.20%	60.42%	60.11%	47.52%	44.25%
AVG100	28.88%	41.08%	32.96%	32.18%	16.96%	21.81%	28.75%	35.21%	32.85%	32.13%	16.96%	21.81%	28.64%	35.28%	32.75%	32.12%	16.96%	21.81%

T403 Query 5

T403 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	81.82%	81.82%	75%	90%	75%	69.23%	75%	75%	75%	90%	75%	69.23%	90%	69.23%	69.23%	90%	75%	69.23%
20%	72%	66.67%	62.07%	64.29%	69.23%	64.29%	60%	62.07%	62.07%	64.29%	69.23%	64.29%	60%	64.29%	64.29%	64.29%	69.23%	64.29%
30%	48.21%	56.25%	48.21%	35.53%	48.21%	44.26%	46.55%	50%	48.21%	35.07%	48.21%	44.26%	34.18%	54%	54%	35.07%	48.21%	44.26%
40%	40.45%	48%	34.29%	30.51%	36.74%	41.38%	32.73%	37.11%	33.96%	30.25%	36.74%	41.38%	28.80%	46.15%	43.90%	30.25%	36.74%	41.38%
50%	35.66%	37.71%	30.67%	28.40%	23.47%	32.39%	31.94%	33.09%	30.46%	28.05%	23%	32.39%	31.94%	28.40%	28.40%	28.05%	23%	32.39%
60%	37.16%	32.16%	22.45%	22%	4.48%	19.71%	26.32%	24.89%	22.09%	21.74%	4.17%	19.57%	20.37%	18.84%	18.58%	21.74%	4.17%	19.57%
70%	23.70%	27.35%	19.05%	17.93%	4.41%	16.50%	20.19%	19.28%	18.82%	17.78%	4.15%	16.37%	17.78%	17.30%	17.20%	17.53%	4.14%	16.37%
80%	21.16%	23.86%	16.30%	11.89%	4.46%	11.46%	14.12%	16.01%	16.12%	11.76%	4.22%	11.39%	11.44%	17.30%	16.78%	11.34%	4.23%	11.39%
90%	12.83%	14.02%	9.64%	8.36%	3.86%	9.59%	10.09%	9.79%	9.51%	8.23%	3.70%	9.47%	7.80%	10.15%	10.06%	7.87%	3.69%	9.47%
100%	7.70%	7.30%	6.16%	3.33%	1.43%	5.34%	5.70%	6.23%	6.02%	3.26%	1.40%	5.17%	2.97%	3.09%	2.97%	2.97%	1.40%	5.17%
Avg30	67.34%	68.25%	61.76%	63.27%	64.15%	59.26%	60.52%	62.36%	61.76%	63.12%	64.15%	59.26%	61.39%	62.51%	62.51%	63.12%	64.15%	59.26%
AVG100	38.07%	39.51%	32.38%	31.22%	27.13%	31.42%	32.26%	33.35%	32.23%	31.04%	26.98%	31.35%	30.53%	32.87%	32.54%	30.91%	26.98%	31.35%

T403 Query 6

T403 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	3.86%	90%	90%	90%	81.82%	100%	100%	90%	90%	90%	81.82%	100%	90%	100%	90%	90%	81.82%	100%
20%	1.38%	51.43%	72%	56.25%	81.82%	54.55%	75%	72%	72%	56.25%	81.82%	54.55%	56.25%	52.94%	56.25%	56.25%	81.82%	54.55%
30%	0.97%	39.13%	67.50%	56.25%	65.85%	55.10%	58.70%	67.50%	67.50%	56.25%	65.85%	55.10%	56.25%	54%	57.45%	56.25%	65.85%	55.10%
40%	0.87%	41.38%	51.43%	44.44%	50.70%	47.37%	50.70%	52.94%	51.43%	44.44%	50.70%	47.37%	46.15%	51.43%	45.57%	44.44%	50.70%	47.37%
50%	0.86%	25.41%	43.81%	37.71%	28.05%	30.26%	43.81%	50%	46.94%	37.71%	28.05%	32.17%	32.62%	43.81%	41.44%	37.71%	28.05%	32.17%
60%	0.71%	19.71%	25%	32.74%	20.76%	25.46%	41.67%	37.42%	34.81%	35.95%	22.54%	30.22%	30.22%	35.48%	33.74%	35.95%	22.54%	30.22%
70%	0.71%	13.42%	16.71%	21.92%	18.77%	17.30%	31.22%	27.59%	25.40%	22.62%	21.12%	18.77%	23.97%	29.36%	25.30%	22.62%	21.12%	18.77%
80%	0%	1.31%	1.79%	18.39%	9.59%	10.61%	23.78%	22.81%	22.32%	21.92%	14.48%	17.59%	21.47%	20.98%	19.73%	21.92%	14.48%	17.59%
90%	0%	0%	0%	0%	0%	0%	14.49%	11.16%	11.88%	13.21%	8.06%	10.51%	15.05%	14.29%	13.90%	13.18%	8.04%	10.51%
100%	0%	0%	0%	0%	0%	0%	7.63%	1.11%	2.03%	6.22%	3.71%	6.64%	4.71%	0.92%	6.72%	6.20%	3.70%	6.64%
Avg30	2.07%	60.19%	76.50%	67.50%	76.50%	69.88%	77.90%	76.50%	76.50%	67.50%	76.50%	69.88%	67.50%	68.98%	67.90%	67.50%	76.50%	69.88%
AVG100	0.94%	28.18%	36.82%	35.77%	35.74%	34.06%	44.70%	43.25%	42.43%	38.46%	37.82%	37.29%	37.67%	40.32%	39.01%	38.45%	37.81%	37.29%

T403 Query 7

T403 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.57%	81.82%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
20%	0.41%	85.71%	81.82%	81.82%	81.82%	64.29%	81.82%	81.82%	81.82%	78.26%	81.82%	64.29%	81.82%	81.82%	81.82%	78.26%	81.82%	64.29%
30%	0.44%	81.82%	71.05%	54%	54%	65.85%	65.85%	71.05%	71.05%	56.25%	52.94%	64.29%	57.45%	56.25%	52.94%	56.25%	52.94%	64.29%
40%	0.40%	53.73%	57.14%	40.91%	40.91%	37.50%	50%	57.14%	56.25%	42.35%	39.56%	48%	36.74%	41.38%	37.90%	42.35%	39.56%	48%
50%	0.38%	43.81%	42.20%	28.75%	24.34%	27.55%	47.92%	46.47%	46.47%	41.82%	27.55%	32.86%	33.33%	31.72%	29.30%	41.82%	27.55%	32.86%
60%	0.45%	35.03%	37.93%	25%	16.08%	17.13%	39.57%	42.31%	42.31%	27.36%	17.97%	19.44%	25.11%	25%	23.21%	27.36%	17.97%	19.37%
70%	0.33%	21.05%	22.62%	12.05%	6.86%	12.12%	28.44%	30.62%	29.91%	19.34%	11.39%	13.39%	17.73%	16.33%	14.95%	19.22%	11.35%	13.36%
80%	0.29%	11.70%	11.70%	9.29%	3.88%	7.05%	17.55%	20.98%	19.73%	11.93%	6.95%	11.61%	10.98%	10.99%	10.21%	11.81%	6.91%	11.57%
90%	0%	0%	0%	0%	0%	0%	11.92%	13.42%	12.33%	8.96%	3.62%	7.21%	7.61%	8.66%	8.10%	8.71%	3.53%	6.78%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.93%	0.31%	0.30%	1.12%	1.78%	0.41%
Avg30	0.48%	83.12%	80.96%	75.27%	75.27%	73.38%	79.22%	80.96%	80.96%	74.84%	74.92%	72.86%	76.42%	76.02%	74.92%	74.84%	74.92%	72.86%
AVG100	0.33%	41.47%	41.45%	34.18%	31.79%	32.15%	43.31%	45.38%	44.99%	37.63%	33.18%	35.11%	36.17%	36.25%	34.87%	37.69%	33.34%	35.09%

T403 Query 8

T403 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.54%	81.82%	75%	81.82%	60%	60%	75%	75%	75%	81.82%	60%	60%	64.29%	81.82%	81.82%	81.82%	60%	60%
20%	0.75%	81.82%	81.82%	60%	43.90%	37.50%	75%	81.82%	81.82%	60%	43.90%	37.50%	69.23%	60%	60%	60%	43.90%	37.50%
30%	0.88%	81.82%	84.38%	57.45%	35.07%	25.23%	81.82%	84.38%	84.38%	56.25%	35.07%	24.32%	55.10%	60%	60%	56.25%	35.07%	24.32%
40%	0.75%	56.25%	67.93%	42.35%	17.31%	29.03%	55.39%	63.16%	67.93%	48.65%	17.23%	27.91%	52.17%	52.17%	50.70%	48.65%	17.23%	27.91%
50%	0.85%	13.41%	37.71%	25.28%	5.56%	26.59%	12.89%	42.59%	41.82%	42.59%	7.99%	25.56%	35.12%	46.47%	44.66%	42.59%	7.99%	25.56%
60%	0.61%	8.86%	26.19%	18.21%	3.43%	26.44%	8.97%	38.19%	37.67%	24.66%	4.79%	24.66%	11.73%	27.09%	24.44%	24.66%	4.79%	24.55%
70%	0%	7.19%	6.62%	8.22%	1.21%	18.82%	8.98%	29.63%	28.57%	22.86%	1.32%	17.83%	10.96%	20.85%	18.66%	22.86%	1.32%	17.78%
80%	0%	0%	0%	0%	1.09%	7.11%	6.87%	18.34%	16.44%	17.30%	1.14%	12.01%	7.47%	16.04%	13.96%	17.22%	1.14%	11.91%
90%	0%	0%	0%	0%	0%	0%	6.13%	6.60%	5.87%	8.66%	1.22%	7.78%	5.71%	8.47%	6.37%	8.52%	1.21%	7.69%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.06%	81.82%	80.40%	66.42%	46.32%	40.91%	77.27%	80.40%	80.40%	66.02%	46.32%	40.61%	62.87%	67.27%	67.27%	66.02%	46.32%	40.61%
AVG100	0.54%	33.12%	37.96%	29.33%	16.76%	23.07%	33.10%	43.97%	43.95%	36.28%	17.27%	23.76%	31.18%	37.29%	36.06%	36.26%	17.26%	23.72%

T403 Query 9

T403 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.99%	64.29%	81.82%	81.82%	81.82%	81.82%	69.23%	75%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	64.29%	81.82%
20%	0.96%	54.55%	66.67%	60%	32.73%	40%	72%	85.71%	78.26%	72%	42.86%	54.55%	54.55%	81.82%	81.82%	85.71%	58.07%	54.55%
30%	0.99%	42.86%	45.76%	45%	35.53%	37.50%	55.10%	61.36%	62.79%	56.25%	31.40%	43.55%	45.76%	50%	48.21%	62.79%	23.08%	40.30%
40%	0.92%	29.75%	32.43%	33.33%	20.93%	23.68%	40%	63.16%	56.25%	43.37%	27.07%	35.29%	37.50%	43.37%	42.86%	47.37%	27.48%	36.74%
50%	0.81%	25.56%	27.71%	28.05%	16.43%	22.77%	40.71%	46.94%	43.40%	34.33%	24.47%	33.82%	33.33%	34.33%	35.39%	37.71%	26.74%	31.94%
60%	0%	12.70%	16.22%	19.03%	10.85%	17.41%	30.73%	41.67%	35.95%	31.25%	21.48%	29.89%	28.65%	31.61%	28.80%	32.16%	19.64%	30.56%
70%	0%	0%	0.98%	7.85%	6.01%	10.60%	26.89%	28.44%	27.23%	23.97%	18.44%	21.84%	23.44%	24.24%	23.62%	23.79%	17.02%	22.86%
80%	0%	0%	0%	0%	0%	0%	16.90%	16.78%	14.37%	13.62%	14.15%	16.67%	12.29%	15.84%	12.72%	13.37%	13.39%	15.24%
90%	0%	0%	0%	0%	0%	0%	12.79%	15.27%	14.75%	11.60%	7.03%	13.60%	9.48%	11.01%	9.66%	9.80%	8.21%	13.02%
100%	0%	0%	0%	0%	0%	0%	3.85%	1.07%	3.91%	3.89%	1.54%	1.17%	3.68%	1.06%	4.17%	3.66%	0.57%	1.16%
Avg30	0.98%	53.90%	64.75%	62.27%	50.02%	53.11%	65.44%	74.03%	74.29%	70.02%	52.02%	59.97%	60.71%	71.21%	70.62%	76.77%	48.48%	58.89%
AVG100	0.47%	22.97%	27.16%	27.51%	20.43%	23.38%	36.82%	43.54%	41.87%	37.21%	27.03%	33.22%	33.05%	37.51%	36.90%	39.82%	25.85%	32.82%

T403 Query 10

T403 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	90%	90%	81.82%	90%	90%	81.82%	81.82%	81.82%	81.82%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	81.82%
20%	90%	85.71%	85.71%	85.71%	85.71%	90%	85.71%	85.71%	85.71%	85.71%	85.71%	90%	90%	90%	85.71%	85.71%	85.71%	90%
30%	65.85%	71.05%	90%	84.38%	84.38%	90%	84.38%	90%	90%	84.38%	84.38%	90%	90%	90%	84.38%	84.38%	84.38%	90%
40%	40.45%	39.13%	75%	80%	80%	80%	81.82%	83.72%	75%	80%	80%	80%	80%	80%	80%	80%	80%	80%
50%	11.41%	10.88%	15.03%	66.67%	70.77%	74.19%	77.97%	58.23%	15.39%	63.89%	67.65%	74.19%	73.02%	74.19%	63.89%	63.89%	67.65%	74.19%
60%	3.46%	3.56%	9.62%	46.61%	50.93%	38.46%	39.57%	14.67%	9.60%	46.61%	49.55%	38.19%	46.61%	38.19%	46.61%	46.61%	49.55%	38.19%
70%	0.92%	0.96%	9.36%	24.81%	19.45%	20.19%	15.76%	8.42%	10.36%	32%	33.86%	20.38%	28.07%	20.38%	31.07%	32%	33.86%	20.38%
80%	0%	0%	7.93%	8.66%	6.36%	7.60%	13.91%	0.35%	8.95%	19.89%	16.22%	11.34%	23.47%	12.21%	18.96%	19.89%	16.12%	11.34%
90%	0%	0%	0%	0%	0%	0%	8.60%	0.39%	8.35%	8.74%	8.05%	6.20%	18.47%	7.64%	8.41%	8.74%	7.96%	6.20%
100%	0%	0%	0%	0%	0%	0%	4.66%	0.43%	4.62%	1.39%	1.40%	0.48%	2.44%	0.55%	1.40%	1.33%	1.30%	0.48%
Avg30	81.95%	82.26%	85.84%	86.70%	86.70%	87.27%	83.97%	85.84%	85.84%	86.70%	86.70%	87.27%	87.27%	87.27%	86.70%	86.70%	86.70%	87.27%
AVG100	30.21%	30.13%	37.45%	48.68%	48.76%	48.23%	49.42%	42.37%	38.98%	51.26%	51.68%	49.26%	53.39%	49.50%	51.04%	51.25%	51.65%	49.26%

T403 Query 1 entropy

T403 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	90%	90%	90%	90%	75%	90%	90%	90%	90%	90%	75%	90%	90%	90%	90%	90%	75%	90%
20%	69.23%	69.23%	75%	81.82%	78.26%	81.82%	72%	78.26%	75%	81.82%	78.26%	81.82%	69.23%	87.71%	81.82%	87.71%	78.26%	81.82%
30%	67.50%	69.23%	65.85%	52.94%	57.45%	54%	58.70%	69.23%	65.85%	52.94%	57.45%	54%	51.92%	77.14%	57.45%	77.14%	57.45%	54%
40%	53.73%	58.07%	60%	40%	34.95%	35.29%	56.25%	66.67%	60%	40%	34.95%	35.29%	32.14%	56.25%	43.37%	40%	34.95%	35.29%
50%	37.40%	33.58%	33.82%	35.12%	23.47%	30.67%	37.10%	42.99%	33.82%	35.12%	23.47%	30.87%	37.10%	30.26%	33.58%	35.12%	23.47%	30.87%
60%	25.35%	26.32%	24.44%	28.65%	7.98%	25.11%	22.45%	28.35%	24.44%	28.50%	7.98%	25.11%	27.09%	27.50%	28.65%	28.50%	7.98%	25.11%
70%	19.88%	19.34%	23.27%	20.71%	7.74%	24.52%	21.84%	25.70%	23.27%	20.65%	7.73%	24.62%	20.19%	22.86%	20.32%	20.65%	7.73%	24.62%
80%	14.57%	14.69%	20.98%	18.11%	6.97%	18.25%	20.80%	23.55%	20.98%	17.76%	6.93%	18.20%	15.08%	20.28%	17.98%	17.76%	6.92%	18.20%
90%	11.73%	10.98%	14.04%	12.89%	4.59%	13.97%	19.43%	16.17%	13.95%	12.71%	4.52%	13.97%	11.20%	13.95%	13.00%	12.69%	4.45%	13.97%
100%	0%	0%	0%	0%	0%	0%	5.76%	0.75%	3.92%	2.97%	0.80%	1.81%	2.40%	0.71%	3.52%	2.77%	0.72%	1.71%
Avg30	75.58%	76.15%	76.95%	74.92%	70.24%	75.27%	73.57%	79.16%	76.95%	74.92%	70.24%	75.27%	70.38%	84.95%	76.42%	84.95%	70.24%	75.27%
AVG100	38.94%	39.14%	40.74%	38.02%	29.64%	37.36%	40.43%	44.17%	41.12%	38.25%	29.71%	37.57%	35.64%	42.67%	38.97%	41.23%	29.69%	37.56%

T403 Query 2 entropy

T403 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	75%	75%	90%	81.82%	81.82%	75%	90%	90%	90%	81.82%	81.82%	75%	81.82%	75%	75%	81.82%	81.82%	75%
20%	72%	75%	66.67%	64.29%	66.67%	66.67%	62.07%	66.67%	66.67%	64.29%	66.67%	66.67%	56.25%	69.23%	69.23%	69.23%	66.67%	66.67%
30%	52.94%	52.94%	56.25%	55.10%	58.70%	64.29%	54%	56.25%	56.25%	55.10%	58.70%	64.29%	56.25%	65.85%	65.85%	65.85%	58.70%	64.29%
40%	37.11%	40%	40.91%	57.14%	53.73%	56.25%	43.37%	41.38%	40.91%	57.14%	53.73%	56.25%	56.25%	56.25%	56.25%	57.14%	53.73%	56.25%
50%	29.30%	16.08%	33.09%	42.20%	35.12%	36.80%	28.75%	32.86%	33.09%	42.20%	35.12%	36.80%	42.20%	41.44%	41.44%	42.20%	35.12%	36.80%
60%	16.37%	14.51%	19.93%	35.03%	29.26%	29.57%	25.94%	20.15%	19.93%	35.03%	29.26%	29.57%	38.19%	31.25%	32.93%	35.03%	29.26%	29.57%
70%	10.27%	11.74%	12.62%	20.51%	18.82%	15.13%	15.53%	12.31%	12.62%	20.38%	18.71%	15.09%	21.55%	15.46%	14.78%	20.38%	18.71%	15.09%
80%	9.25%	7.11%	8.80%	13.85%	7.95%	9.37%	6.84%	9.01%	8.77%	13.67%	7.86%	9.25%	13.96%	10.47%	9.80%	13.65%	7.86%	9.25%
90%	3.57%	2.96%	5.35%	7.50%	2.66%	7.71%	4.53%	5.59%	5.21%	7.34%	2.60%	7.61%	6.43%	6.33%	6.41%	7.33%	2.59%	7.61%
100%	1.42%	1.43%	1.80%	1.05%	1.23%	2.68%	1.53%	2.04%	1.56%	0.98%	1.11%	2.47%	0.72%	1.90%	1.50%	0.70%	1.03%	2.46%
Avg30	66.65%	67.65%	70.97%	67.07%	69.06%	68.65%	68.69%	70.97%	70.97%	67.07%	69.06%	68.65%	64.77%	70.03%	70.03%	72.30%	69.06%	68.65%
AVG100	30.72%	29.68%	33.54%	37.85%	35.59%	36.35%	33.26%	33.63%	33.50%	37.80%	35.56%	36.30%	37.36%	37.32%	37.32%	39.33%	35.55%	36.30%

T403 Query 3 entropy

T403 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	69.23%	69.23%	64.29%	75%	69.23%	75%	69.23%	69.23%	64.29%	75%	69.23%	75%	75%	90%	75%	91%	69.23%	75%
20%	46.15%	34.62%	45%	81.82%	66.67%	75%	40.91%	60%	45%	81.82%	66.67%	75%	81.82%	81.82%	81.82%	81.82%	66.67%	75%
30%	29.03%	33.33%	50.94%	69.23%	61.36%	71.05%	47.37%	58.70%	50.94%	69.23%	61.36%	71.05%	62.79%	77.14%	72.97%	77.14%	61.36%	71.05%
40%	23.23%	31.30%	40.45%	33.33%	48%	40%	46.15%	43.37%	40%	31.03%	47.37%	40%	31.30%	38.30%	31.30%	31.03%	47.37%	40%
50%	18.55%	21.50%	33.09%	24.60%	24.60%	35.12%	40%	33.33%	32.62%	22.44%	23.83%	34.85%	21.80%	30.07%	22.12%	22.44%	23.83%	34.85%
60%	19.10%	20.07%	28.21%	25.46%	16.98%	28.35%	29.89%	27.92%	27.64%	22.73%	16.32%	27.78%	23.01%	27.50%	22.36%	22.73%	16.32%	27.78%
70%	19.05%	18.61%	22.78%	25.30%	10.08%	26.45%	24.71%	19.51%	21.92%	22.62%	9.76%	25.81%	23.11%	24.71%	21.26%	22.62%	9.76%	25.81%
80%	14.01%	14.15%	18.86%	23.63%	9.57%	21.41%	21.22%	14.31%	18.11%	20.74%	9.23%	20.92%	20.56%	18.48%	18.81%	20.62%	9.22%	20.92%
90%	7.43%	6.29%	5.22%	21.69%	9.77%	15.27%	10.15%	6.29%	3.21%	18.68%	9.35%	14.91%	16.21%	14.67%	16.87%	18.59%	9.32%	14.91%
100%	0%	0%	0%	0%	0%	0%	0.75%	3.10%	0.74%	5.83%	4.11%	3.64%	4.58%	2.97%	6.72%	4.68%	3.41%	3.30%
Avg30	48.14%	45.73%	53.41%	75.35%	65.75%	73.68%	52.50%	62.64%	53.41%	75.35%	65.75%	73.68%	73.20%	82.99%	76.60%	83.32%	65.75%	73.68%
AVG100	24.58%	24.91%	30.88%	38.01%	31.63%	38.76%	33.04%	33.58%	30.45%	37.01%	31.72%	38.90%	36.02%	40.57%	36.92%	39.27%	31.65%	38.86%

T403 Query 4 entropy

T403 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	75%	75%	81.82%	81.82%	60%	60%	75%	81.82%	81.82%	81.82%	60%	60%	75%	81.82%	81.82%	81.82%	60%	60%
20%	72%	78.26%	78.26%	78.26%	48.65%	51.43%	62.07%	78.26%	78.26%	78.26%	48.65%	51.43%	60%	78.26%	78.26%	78.26%	48.65%	51.43%
30%	62.79%	75%	60%	57.45%	44.26%	39.71%	50%	71.05%	60%	57.45%	44.26%	39.71%	51.92%	51.92%	58.70%	71.05%	44.26%	39.71%
40%	56.25%	65.46%	55.39%	55.39%	21.56%	32.43%	43.90%	56.25%	54.55%	55.39%	21.56%	32.43%	41.38%	53.73%	55.39%	55.39%	21.56%	32.43%
50%	54.12%	55.42%	51.69%	46%	5.64%	33.82%	42.59%	53.49%	51.11%	45.55%	5.64%	33.82%	41.07%	51.69%	46.94%	45.55%	5.64%	33.82%
60%	41.99%	52.89%	33.13%	32.54%	6.14%	22.82%	32.35%	35.48%	32.74%	32.35%	6.14%	22.82%	31.61%	33.54%	33.54%	32.35%	6.14%	22.82%
70%	34.41%	34.04%	22.30%	23.44%	6.64%	19.45%	22.86%	23.88%	21.84%	23.19%	6.64%	19.45%	22.46%	24.24%	23.97%	23.19%	6.64%	19.45%
80%	20.39%	25.26%	17.38%	15.50%	2.09%	10.43%	13.93%	19.01%	16.98%	15.18%	2.09%	10.41%	15.77%	18.30%	15.80%	15.15%	2.09%	10.41%
90%	17.23%	13.53%	9.99%	12.64%	2.17%	8.80%	11.08%	11.25%	9.66%	12.33%	2.16%	8.79%	12.13%	16.08%	13.06%	12.28%	2.16%	8.79%
100%	2.23%	2.53%	2.50%	2.50%	1.76%	2.09%	2.20%	2.51%	2.46%	2.47%	1.75%	2.06%	2.18%	2.51%	2.47%	2.45%	1.75%	2.06%
Avg30	69.93%	76.09%	73.36%	72.51%	50.97%	50.38%	62.36%	77.04%	73.36%	72.51%	50.97%	50.38%	62.31%	70.67%	72.93%	77.04%	50.97%	50.38%
AVG100	43.64%	47.74%	41.25%	40.55%	19.89%	28.10%	35.60%	43.30%	40.94%	40.40%	19.89%	28.09%	35.35%	41.21%	40.99%	41.75%	19.89%	28.09%

T403 Query 5 entropy

T403 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	81.82%	81.82%	69.23%	90%	75%	81.82%	75%	81.82%	69.23%	90%	75%	81.82%	90%	90%	90%	90%	75%	81.82%
20%	75%	75%	69.23%	69.23%	81.82%	56.25%	69.23%	62.07%	69.23%	69.23%	81.82%	56.25%	69.23%	72%	69.23%	69.23%	81.82%	56.25%
30%	60%	65.85%	56.25%	50%	54%	45.76%	50%	56.25%	56.25%	49.09%	54%	45.76%	47.37%	52.94%	50.94%	49.09%	54%	45.76%
40%	43.37%	52.94%	45%	37.90%	39.13%	39.13%	36.36%	46.15%	44.44%	37.50%	39.13%	39.13%	38.30%	34.29%	32.73%	37.50%	39.13%	39.13%
50%	42.59%	47.92%	41.44%	35.66%	4.24%	33.09%	35.66%	38.98%	41.07%	34.85%	4.00%	33.09%	36.51%	26.29%	25.56%	34.85%	4.00%	33.09%
60%	40.15%	34.38%	35.03%	28.95%	4.41%	16.04%	29.57%	33.74%	34.16%	28.50%	4.11%	16.04%	25.58%	25.11%	24.34%	28.50%	4.11%	16.04%
70%	27.95%	29.22%	23.53%	21.19%	4.43%	14.38%	22.07%	25.20%	23.19%	20.98%	4.17%	14.13%	19.75%	19.45%	19.39%	20.98%	4.17%	14.13%
80%	25.09%	29.44%	16.86%	15.50%	4.44%	11.97%	17.51%	19.52%	16.59%	15.27%	4.20%	11.77%	15.53%	16.40%	16.19%	15.27%	4.20%	11.77%
90%	14.86%	16.40%	11.28%	11.45%	2.95%	8.95%	11.30%	11.62%	11.16%	11.30%	2.85%	8.77%	10.58%	12.33%	12.17%	11.22%	2.85%	8.77%
100%	9.89%	6.78%	8.10%	3.22%	1.60%	4.90%	7.23%	8.35%	7.91%	3.13%	1.57%	4.70%	2.90%	6.01%	5.91%	2.97%	1.56%	4.70%
Avg30	72.27%	74.22%	64.90%	69.74%	70.27%	61.28%	64.74%	66.71%	64.90%	69.44%	70.27%	61.28%	68.87%	71.65%	70.06%	69.44%	70.27%	61.28%
AVG100	42.07%	43.97%	37.60%	36.31%	27.20%	31.23%	35.39%	38.37%	37.32%	35.98%	27.08%	31.15%	35.58%	35.48%	34.64%	35.96%	27.08%	31.15%

T403 Query 6 entropy

T403 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	3.86%	64.29%	90%	81.82%	81.82%	81.82%	90%	90%	90%	69.23%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%
20%	1.38%	62.07%	75%	58.07%	81.82%	54.55%	75%	81.82%	75%	72%	81.82%	54.55%	58.07%	81.82%	58.07%	81.82%	81.82%	54.55%
30%	0.97%	52.94%	64.29%	52.94%	67.50%	55.10%	65.85%	64.29%	64.29%	62.79%	67.50%	55.10%	54%	67.50%	56.25%	62.79%	67.50%	55.10%
40%	0.87%	50.70%	55.39%	48%	43.37%	43.90%	52.17%	58.07%	55.39%	56.25%	43.37%	43.90%	46.75%	57.14%	52.94%	56.25%	43.37%	43.90%
50%	0.86%	41.82%	41.44%	40.71%	27.71%	31.08%	43.40%	51.69%	52.27%	42.99%	27.71%	31.94%	35.39%	45.55%	48.42%	42.99%	27.71%	31.94%
60%	0.71%	31.98%	28.21%	32.74%	23.81%	23.21%	45.83%	43.31%	35.71%	43.65%	24.34%	25.82%	33.95%	43.65%	41.99%	43.65%	24.34%	25.82%
70%	0.71%	24.06%	16.93%	23.70%	20.71%	17.11%	28.96%	32.82%	30.77%	30.92%	23.11%	22.07%	24.52%	32.82%	30.62%	30.92%	23.11%	22.07%
80%	0%	5.92%	1.26%	22.60%	9.43%	14.01%	25.17%	27.34%	23.63%	21.47%	19.11%	18.20%	23.70%	21.60%	20.86%	21.28%	19.11%	18.20%
90%	0%	0%	0%	0%	0%	0%	15.16%	16.70%	13.85%	19.34%	8.57%	13.60%	15.92%	19.86%	16.84%	19.20%	8.54%	13.60%
100%	0%	0%	0%	0%	0%	0%	5.10%	1.04%	1.40%	7.55%	3.01%	5.74%	5.47%	0.88%	6.36%	7.26%	2.99%	5.73%
Avg30	2.07%	59.77%	76.43%	64.27%	77.05%	63.82%	76.95%	78.70%	76.43%	68.01%	77.05%	63.82%	64.63%	77.05%	65.38%	75.48%	77.05%	63.82%
AVG100	0.94%	33.38%	37.25%	36.06%	35.62%	32.08%	44.66%	46.71%	44.23%	42.62%	38.03%	35.27%	37.96%	45.26%	41.42%	44.80%	38.03%	35.27%

T403 Query 7 entropy

T403 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.57%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
20%	0.41%	78.26%	78.26%	78.26%	78.26%	64.29%	75%	78.26%	78.26%	75%	78.26%	64.29%	78.26%	78.26%	78.26%	75%	78.26%	64.29%
30%	0.44%	71.05%	77.14%	61.36%	60%	62.79%	75%	77.14%	77.14%	64.29%	60%	61.36%	57.45%	60%	57.45%	64.29%	60%	61.36%
40%	0.40%	62.07%	52.94%	44.44%	36.74%	45%	63.16%	52.94%	52.17%	50.70%	35.64%	45%	41.86%	45.57%	45%	50.70%	35.64%	45%
50%	0.38%	50.55%	48.42%	41.07%	32.62%	35.39%	43.40%	47.92%	47.92%	43.81%	31.72%	35.94%	37.71%	46%	41.44%	43.81%	31.72%	35.94%
60%	0.45%	36.42%	32.35%	30.39%	15.03%	23.91%	37.42%	39.01%	38.73%	31.43%	23.31%	21.91%	29.57%	33.33%	35.95%	31.43%	23.31%	21.83%
70%	0.33%	25.81%	23.36%	20%	8.13%	16.50%	33.68%	31.37%	30.77%	20.71%	10.63%	16.93%	24.62%	25.20%	19.10%	20.65%	10.60%	16.89%
80%	0.29%	15.43%	10.18%	13.93%	4.23%	8.74%	21.86%	22.88%	21.73%	14.81%	5.36%	15.37%	14.84%	17.22%	15.60%	14.78%	5.35%	15.34%
90%	0%	0%	0%	0%	0%	0%	14.19%	16.80%	16.02%	11.60%	4.21%	8.77%	11.92%	14.39%	12.11%	11.53%	4.20%	8.72%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.92%	0.30%	0.32%	1.19%	1.82%	0.44%
Avg30	0.48%	79.77%	81.80%	76.54%	76.09%	72.36%	80.00%	81.80%	81.80%	76.43%	76.09%	71.88%	75.24%	76.09%	75.24%	76.43%	76.09%	71.88%
AVG100	0.33%	42.96%	41.27%	37.95%	32.50%	34.66%	45.37%	45.63%	45.27%	40.23%	33.91%	35.96%	38.71%	41.03%	39.52%	40.34%	34.09%	35.98%

T403 Query 8 entropy

T403 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.54%	81.82%	81.82%	81.82%	60%	64.29%	75%	81.82%	81.82%	81.82%	60%	60%	81.82%	81.82%	81.82%	81.82%	60%	60%
20%	0.75%	85.71%	85.71%	75%	43.90%	33.33%	81.82%	81.82%	85.71%	75%	43.90%	42.86%	75%	72%	72%	75%	43.90%	42.86%
30%	0.88%	84.38%	77.14%	61.36%	33.33%	30.68%	81.82%	84.38%	77.14%	67.50%	33.33%	23.68%	60%	64.29%	64.29%	67.50%	33.33%	23.68%
40%	0.75%	63.16%	66.67%	46.75%	26.28%	26.67%	63.16%	60%	66.67%	63.16%	26.09%	28.80%	59.02%	59.02%	58.07%	63.16%	26.09%	28.80%
50%	0.85%	43.40%	44.66%	32.62%	15.44%	23.71%	38.66%	52.87%	44.66%	46.94%	14.94%	28.05%	46.47%	33.09%	31.08%	46.94%	14.94%	28.05%
60%	0.61%	14.40%	30.06%	11.91%	15.94%	21.83%	32.35%	15.24%	43.31%	32.93%	15.45%	25.58%	29.10%	34.59%	32.74%	32.93%	15.45%	25.46%
70%	0%	8.23%	13.53%	10.13%	11.29%	17.44%	31.22%	9.89%	32.32%	13.09%	12.19%	19.81%	14.41%	28.07%	26.12%	13.04%	12.14%	19.63%
80%	0%	0%	0%	0%	1.17%	7.47%	9.51%	9.44%	21.66%	12.99%	2.18%	13.11%	12.35%	21.41%	18.86%	12.90%	2.17%	12.83%
90%	0%	0%	0%	0%	0%	0%	7.09%	8.18%	13.31%	11.08%	1.29%	9.32%	10.50%	17.71%	15.19%	10.85%	1.29%	9.18%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.06%	83.97%	81.56%	72.73%	45.75%	42.77%	79.55%	82.67%	81.56%	74.77%	45.75%	42.18%	72.27%	72.70%	72.70%	74.77%	45.75%	42.18%
AVG100	0.54%	38.11%	39.96%	31.96%	20.73%	22.54%	42.06%	40.36%	46.66%	40.45%	20.94%	25.12%	38.87%	41.20%	40.02%	40.41%	20.93%	25.05%

T403 Query 9 entropy

T403 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.99%	60%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	60%	81.82%
20%	0.96%	64.29%	78.26%	90%	42.86%	64.29%	75%	81.82%	85.71%	85.71%	40.91%	78.26%	81.82%	81.82%	85.71%	85.71%	60%	64.29%
30%	0.99%	54%	61.36%	58.70%	30%	40.91%	64.29%	71.05%	64.29%	56.25%	32.53%	43.55%	62.79%	67.50%	55.10%	61.36%	24.55%	52.94%
40%	0.92%	40%	41.86%	53.73%	25.71%	33.03%	54.55%	67.93%	61.02%	52.94%	28.35%	41.86%	48.65%	52.17%	46.75%	57.14%	26.09%	37.50%
50%	0.81%	32.17%	27.71%	41.44%	23.47%	30.07%	47.42%	58.23%	48.42%	45.55%	25.41%	38.33%	41.44%	45.10%	38.33%	45.55%	25.28%	35.66%
60%	0%	15.85%	16.67%	27.09%	17.24%	21.91%	40.74%	43.31%	42.31%	40.44%	22%	36.42%	36.42%	36.91%	33.13%	34.59%	19.64%	33.54%
70%	0%	0%	9.01%	12.85%	6.31%	11.27%	30.77%	32.49%	34.41%	30.19%	18.71%	25.20%	29.22%	27.95%	27.00%	29.77%	19.05%	25.20%
80%	0%	0%	0%	0%	0%	0%	19.62%	21.66%	20.06%	18.39%	13.37%	18.34%	16.48%	19.21%	16.90%	17.72%	14.23%	16.78%
90%	0%	0%	0%	0%	0%	0%	17.05%	17.16%	15.86%	13.42%	5.88%	15.74%	12.48%	14.75%	13.29%	13.47%	8.78%	15.24%
100%	0%	0%	0%	0%	0%	0%	3.85%	1.07%	3.68%	3.98%	1.34%	1.18%	3.72%	1.06%	4.24%	3.70%	0.49%	1.16%
Avg30	0.98%	59.43%	73.81%	76.84%	51.56%	62.34%	73.70%	78.23%	77.27%	74.59%	51.75%	67.88%	75.48%	77.05%	74.21%	76.30%	48.18%	66.35%
AVG100	0.47%	26.63%	31.67%	36.56%	22.74%	28.33%	43.51%	47.65%	45.76%	42.87%	27.03%	38.07%	41.48%	42.83%	40.23%	43.08%	25.81%	36.41%

T403 Query 10 entropy

T403 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	90%	100%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	90%	90%	90%	90%	90%	90%
20%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
30%	72.97%	62.79%	90%	84.38%	84.38%	84.38%	84.38%	90%	90%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%
40%	47.37%	40%	76.60%	78.26%	80%	78.26%	81.82%	83.72%	76.60%	78.26%	80%	78.26%	80%	78.26%	78.26%	78.26%	80%	78.26%
50%	11.33%	7.64%	14.60%	66.67%	69.70%	69.70%	76.67%	14.98%	15.08%	63.89%	66.67%	69.70%	63.89%	69.70%	63.89%	63.89%	66.67%	69.70%
60%	4.75%	4.29%	9.98%	45.08%	48.67%	44.36%	41.67%	16.22%	10%	50%	51.40%	43.31%	49.55%	43.31%	50%	50%	51.40%	43.31%
70%	0.97%	1%	9.70%	28.83%	21.41%	17.16%	16.04%	8.91%	9.83%	35.17%	33.86%	28.83%	32.65%	28.83%	35.17%	35.17%	33.86%	28.83%
80%	0%	0%	8.43%	8.96%	7.45%	5.18%	14.20%	0.35%	9.38%	25.09%	20.33%	9.35%	23.70%	9.19%	24.01%	25.09%	20.11%	9.35%
90%	0%	0%	0%	0%	0%	0%	9.49%	0.39%	9.11%	10.50%	9.80%	5.26%	10.62%	5.28%	9.56%	10.50%	9.66%	5.26%
100%	0%	0%	0%	0%	0%	0%	4.15%	0.43%	3.92%	1.58%	1.57%	0.67%	1.45%	0.72%	1.59%	1.49%	1.44%	0.67%
Avg30	82.90%	82.84%	85.84%	86.70%	86.70%	86.70%	83.97%	85.84%	85.84%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%
AVG100	31.31%	30.14%	37.68%	48.79%	48.73%	47.47%	49.59%	38.25%	39.15%	52.46%	52.37%	49.55%	52.19%	49.54%	52.26%	52.45%	52.32%	49.55%

T403 Query 1 idf

T403 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	90%	90%	90%	90%	75%	90%	90%	90%	90%	90%	75%	90%	90%	90%	90%	90%	75%	90%
20%	72%	69.23%	75%	81.82%	78.26%	78.26%	72%	75%	75%	81.82%	78.26%	78.26%	60%	81.82%	81.82%	81.82%	78.26%	78.26%
30%	67.50%	69.23%	65.85%	50.94%	56.25%	54%	57.45%	71.05%	65.85%	50.94%	56.25%	54%	56.25%	71.05%	55.10%	69.23%	56.25%	54%
40%	54.55%	58.07%	57.14%	39.56%	40.91%	34.95%	55.39%	56.25%	57.14%	39.56%	40.91%	34.95%	48.65%	54.55%	43.37%	39.56%	40.91%	34.95%
50%	38.66%	33.09%	36.51%	35.39%	24.34%	30.46%	37.10%	42.20%	36.51%	35.39%	24.34%	30.46%	44.23%	29.49%	33.58%	35.39%	24.34%	30.46%
60%	25.58%	26.70%	24.55%	28.35%	8.08%	25.58%	22.27%	29.89%	24.55%	28.21%	8.08%	25.58%	34.16%	29.41%	28.65%	28.21%	8.08%	25.58%
70%	19.28%	19.05%	23.27%	20.71%	7.66%	23.97%	21.70%	25.20%	23.27%	20.58%	7.64%	23.97%	22.15%	22.46%	20.13%	20.58%	7.64%	23.97%
80%	14.51%	14.48%	20.98%	17.72%	6.93%	18.25%	20.62%	22.74%	20.98%	17.38%	6.89%	18.20%	14.66%	20.74%	17.68%	17.38%	6.87%	18.20%
90%	11.57%	10.71%	14.62%	12.46%	4.72%	13.67%	19.03%	15.77%	14.51%	12.29%	4.64%	13.64%	11.42%	13.00%	12.56%	12.28%	4.58%	13.64%
100%	0%	0%	0%	0%	0%	0%	6.03%	0.73%	4.37%	2.92%	0.81%	1.80%	3.55%	0.70%	3.43%	2.74%	0.74%	1.70%
Avg30	76.50%	76.15%	76.95%	74.25%	69.84%	74.09%	73.15%	78.68%	76.95%	74.25%	69.84%	74.09%	68.75%	80.96%	75.64%	80.35%	69.84%	74.09%
AVG100	39.36%	39.06%	40.79%	37.69%	30.21%	36.91%	40.16%	42.88%	41.22%	37.91%	30.28%	37.09%	38.51%	41.32%	38.63%	39.72%	30.27%	37.08%

T403 Query 2 idf

T403 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	75%	75%	90%	81.82%	81.82%	81.82%	81.82%	90%	90%	81.82%	81.82%	81.82%	81.82%	81.82%	75%	81.82%	81.82%	81.82%
20%	72%	75%	64.29%	64.29%	66.67%	64.29%	62.07%	66.67%	64.29%	64.29%	66.67%	64.29%	56.25%	69.23%	69.23%	74.29%	66.67%	64.29%
30%	51.92%	52.94%	56.25%	55.10%	58.70%	58.70%	55.10%	56.25%	56.25%	55.10%	58.70%	58.70%	56.25%	65.85%	65.85%	65.10%	58.70%	58.70%
40%	36%	40%	40.45%	58.07%	55.39%	50%	43.37%	41.38%	40.45%	58.07%	55.39%	50%	57.14%	57.14%	56.25%	58.07%	55.39%	50%
50%	29.11%	16.20%	33.09%	42.20%	34.59%	44.66%	28.93%	32.62%	33.09%	42.20%	34.59%	44.66%	42.20%	41.07%	41.44%	42.20%	34.59%	44.66%
60%	15.94%	14.75%	19.86%	34.81%	30.22%	34.38%	25.46%	20%	19.86%	34.81%	30.22%	34.38%	38.73%	30.73%	32.16%	34.81%	30.22%	34.38%
70%	10.36%	11.51%	12.75%	20.98%	18.34%	15.31%	15.27%	13.06%	12.75%	20.85%	18.23%	15.27%	21.26%	15.50%	14.95%	20.85%	18.23%	15.27%
80%	9.04%	7.11%	8.73%	13.88%	7.76%	11.81%	6.80%	8.98%	8.72%	13.70%	7.67%	11.76%	13.37%	10.46%	9.46%	13.67%	7.67%	11.76%
90%	3.40%	2.96%	5.29%	7.50%	2.71%	8.15%	4.44%	5.60%	5.14%	7.34%	2.64%	8.07%	6.39%	6.34%	6.24%	7.33%	2.64%	8.07%
100%	1.42%	1.43%	1.80%	1.05%	1.23%	3.44%	1.53%	2.04%	1.57%	0.99%	1.11%	3.38%	0.72%	1.90%	1.50%	0.71%	1.03%	3.38%
Avg30	66.31%	67.65%	70.18%	67.07%	69.06%	68.27%	66.33%	70.97%	70.18%	67.07%	69.06%	68.27%	64.77%	72.30%	70.03%	73.74%	69.06%	68.27%
AVG100	30.42%	29.69%	33.25%	37.97%	35.74%	37.25%	32.48%	33.66%	33.21%	37.92%	35.70%	37.23%	37.41%	38.00%	37.21%	39.88%	35.69%	37.23%

T403 Query 3 idf

T403 Q3	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	69.23%	69.23%	64.29%	75%	69.23%	75%	69.23%	69.23%	64.29%	75%	69.23%	75%	75%	90%	75%	75%	69.23%	75%
20%	47.37%	35.29%	46.15%	78.26%	69.23%	78.26%	40.91%	64.29%	46.15%	78.26%	69.23%	78.26%	81.82%	78.26%	78.26%	78.26%	69.23%	78.26%
30%	29.67%	33.33%	50%	71.05%	61.36%	71.05%	47.37%	55.10%	50%	71.05%	61.36%	71.05%	62.79%	77.14%	71.05%	71.05%	61.36%	71.05%
40%	22.36%	31.86%	40.91%	31.58%	50%	40%	40.91%	40.91%	40.45%	29.75%	49.32%	40%	31.58%	38.30%	31.03%	29.75%	49.32%	40%
50%	18.78%	21.70%	34.07%	24.34%	25.14%	34.59%	35.66%	32.86%	33.58%	22.01%	24.34%	34.33%	21.50%	29.87%	22.33%	22.01%	24.34%	34.33%
60%	19.16%	20.45%	28.35%	25.35%	17.92%	28.35%	30.06%	28.06%	27.78%	22.45%	17.24%	27.78%	22.92%	27.36%	22.36%	22.45%	17.24%	27.78%
70%	19.28%	19.69%	23.53%	25.50%	10.36%	26.23%	24.62%	19.28%	22.62%	22.78%	10.02%	25.60%	22.94%	24.52%	21.05%	22.78%	10.02%	25.60%
80%	13.88%	14.04%	18.96%	26.17%	9.59%	21.22%	21.92%	13.99%	18.25%	23.25%	9.28%	20.80%	20.22%	18.43%	19.16%	23.18%	9.34%	20.80%
90%	7.59%	6.34%	5.72%	20.10%	9.70%	15.68%	4.95%	6.43%	3.54%	17.05%	9.30%	15.30%	16.27%	15.16%	16.43%	16.87%	9.31%	15.30%
100%	0%	0%	0%	0%	0%	0%	0.74%	2.65%	0.74%	5.75%	4.11%	3.65%	4.56%	2.98%	6.70%	4.62%	3.42%	3.33%
Avg30	48.76%	45.95%	53.48%	74.77%	66.61%	74.77%	52.50%	62.87%	53.48%	74.77%	66.61%	74.77%	73.20%	81.80%	74.77%	74.77%	66.61%	74.77%
AVG100	24.73%	25.19%	31.20%	37.73%	32.25%	39.04%	31.64%	33.28%	30.74%	36.73%	32.34%	39.18%	35.96%	40.20%	36.34%	36.60%	32.28%	39.14%

T403 Query 4 idf

T403 Q4	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	69.23%	75%	81.82%	81.82%	64.29%	60%	75%	81.82%	81.82%	81.82%	64.29%	60%	75%	81.82%	81.82%	81.82%	64.29%	60%
20%	69.23%	78.26%	78.26%	78.26%	48.65%	50%	60%	78.26%	78.26%	78.26%	48.65%	50%	58.07%	78.26%	78.26%	78.26%	48.65%	50%
30%	61.36%	75%	58.70%	57.45%	47.37%	39.71%	49.09%	67.50%	58.70%	57.45%	47.37%	39.71%	50%	75%	58.70%	57.45%	47.37%	39.71%
40%	57.14%	66.67%	57.14%	56.25%	15.32%	32.43%	43.90%	56.25%	57.14%	56.25%	15.32%	32.43%	41.86%	53.73%	55.39%	56.25%	15.32%	32.43%
50%	52.87%	54.12%	48.94%	45.10%	5.61%	33.82%	43.40%	52.87%	48.42%	44.66%	5.61%	33.82%	40%	50.55%	46%	44.66%	5.61%	33.82%
60%	39.29%	51.40%	34.16%	32.35%	6.18%	22.73%	31.98%	35.03%	33.95%	32.16%	6.18%	22.73%	31.07%	33.13%	33.13%	32.16%	6.18%	22.73%
70%	33.16%	31.84%	22.15%	23.11%	6.23%	18.77%	22.78%	23.02%	21.77%	22.86%	6.23%	18.77%	22.46%	23.97%	23.44%	22.86%	6.23%	18.77%
80%	20.06%	24.92%	16.67%	15.11%	2.09%	10.34%	13.65%	18.77%	16.30%	14.81%	2.09%	10.34%	15.18%	17.46%	15.63%	14.75%	2.08%	10.34%
90%	16.34%	13.12%	9.68%	12.87%	2.15%	8.81%	10.99%	10.59%	9.36%	12.56%	2.15%	8.81%	11.80%	15.71%	13.62%	12.50%	2.14%	8.81%
100%	2.24%	2.51%	2.50%	2.49%	1.98%	2.09%	2.20%	2.50%	2.46%	2.46%	1.96%	2.06%	2.18%	2.50%	2.46%	2.44%	1.95%	2.06%
Avg30	66.61%	76.09%	72.93%	72.51%	53.43%	49.90%	61.36%	75.86%	72.93%	72.51%	53.43%	49.90%	61.02%	78.36%	72.93%	72.51%	53.43%	49.90%
AVG100	42.09%	47.28%	41.00%	40.48%	19.99%	27.87%	35.30%	42.66%	40.82%	40.33%	19.98%	27.87%	34.76%	43.21%	40.84%	40.31%	19.98%	27.87%

T403 Query 5 idf

T403 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	81.82%	81.82%	81.82%	90%	75%	81.82%	75%	81.82%	81.82%	90%	75%	81.82%	90%	81.82%	90%	90%	75%	81.82%
20%	78.26%	78.26%	58.07%	69.23%	81.82%	56.25%	69.23%	62.07%	58.07%	69.23%	81.82%	56.25%	69.23%	78.26%	66.67%	69.23%	81.82%	56.25%
30%	58.70%	64.29%	57.45%	50%	50%	45.76%	50%	58.70%	57.45%	49.09%	50%	45.76%	47.37%	64.29%	49.09%	49.09%	50%	45.76%
40%	43.37%	52.17%	46.15%	36.36%	37.90%	39.56%	37.11%	45.57%	45.57%	36%	37.90%	39.56%	38.71%	33.96%	33.03%	36%	37.90%	39.56%
50%	41.82%	47.92%	38.66%	36.22%	5.08%	33.33%	34.59%	38.33%	37.71%	35.39%	4.92%	33.33%	38.33%	26.44%	25.70%	35.39%	4.95%	33.33%
60%	40.15%	33.74%	33.13%	28.35%	4.43%	16.13%	29.26%	33.33%	32.35%	27.92%	4.12%	16.08%	27.23%	24.89%	24.44%	27.92%	4.12%	16.08%
70%	27.59%	28.83%	24.43%	20.58%	4.47%	14.55%	21.26%	25.10%	23.97%	20.38%	4.20%	14.29%	18.88%	19.16%	19.10%	20.38%	4.20%	14.29%
80%	25%	28.08%	18.81%	15.34%	4.51%	11.89%	16.98%	19.47%	18.53%	15.11%	4.27%	11.70%	14.96%	15.84%	15.73%	15.11%	4.27%	11.70%
90%	14.57%	15.83%	11.33%	11.14%	3.26%	8.92%	10.69%	11.39%	11.20%	10.99%	3.14%	8.74%	10.38%	12.08%	11.75%	10.92%	3.14%	8.74%
100%	9.77%	6.95%	7.98%	3.24%	1.59%	4.87%	7.14%	8.18%	7.79%	3.14%	1.56%	4.66%	2.85%	6.03%	5.94%	2.99%	1.55%	4.66%
Avg30	72.93%	74.79%	65.78%	69.74%	68.94%	61.28%	64.74%	67.53%	65.78%	69.44%	68.94%	61.28%	68.87%	74.79%	68.59%	69.44%	68.94%	61.28%
AVG100	42.10%	43.79%	37.78%	36.05%	26.80%	31.31%	35.13%	38.40%	37.44%	35.73%	26.69%	31.22%	35.79%	36.28%	34.15%	35.70%	26.69%	31.22%

T403 Query 6 idf

T403 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	3.86%	64.29%	90%	81.82%	81.82%	81.82%	90%	90%	90%	69.23%	81.82%	81.82%	81.82%	81.82%	69.23%	69.23%	81.82%	81.82%
20%	1.38%	60%	75%	58.07%	81.82%	54.55%	75%	78.26%	75%	66.67%	81.82%	54.55%	54.55%	56.25%	66.67%	66.67%	81.82%	54.55%
30%	0.97%	51.92%	64.29%	52.94%	67.50%	58.70%	65.85%	64.29%	64.29%	60%	67.50%	58.70%	57.45%	57.45%	60%	60%	67.50%	58.70%
40%	0.87%	50%	55.39%	45.57%	44.44%	43.90%	50%	58.07%	55.39%	51.43%	44.44%	43.90%	44.44%	52.94%	53.73%	51.43%	44.44%	43.90%
50%	0.86%	41.82%	41.82%	40%	27.88%	30.87%	44.66%	52.27%	52.87%	42.99%	27.88%	31.08%	32.86%	50%	40%	42.99%	27.88%	31.08%
60%	0.71%	31.07%	27.92%	32.16%	22.54%	22.92%	44.72%	43.31%	35.95%	45.08%	24.44%	25.46%	27.36%	42.97%	42.31%	45.08%	24.44%	25.46%
70%	0.71%	23.02%	17.34%	23.53%	20.78%	17.07%	34.41%	32.49%	30.62%	33.51%	22.94%	22.07%	25%	32.82%	30.19%	33.51%	22.94%	22.07%
80%	0%	5.89%	1.33%	22.39%	9.11%	13.65%	24.33%	26.64%	23.25%	21.04%	18.86%	18.11%	23.18%	21.22%	21.86%	20.86%	18.86%	18.11%
90%	0%	0%	0%	0%	0%	0%	16.24%	15.89%	13.67%	18.51%	8.69%	14.16%	17.23%	17.67%	17.75%	18.39%	8.66%	14.16%
100%	0%	0%	0%	0%	0%	0%	6.21%	1.05%	1.49%	7.52%	3.13%	5.88%	4.29%	0.90%	10.01%	7.24%	3.12%	5.87%
Avg30	2.07%	58.74%	76.43%	64.27%	77.05%	65.02%	76.95%	77.52%	76.43%	65.30%	77.05%	65.02%	64.60%	65.17%	65.30%	65.30%	77.05%	65.02%
AVG100	0.94%	32.80%	37.31%	35.65%	35.59%	32.35%	45.14%	46.23%	44.25%	41.60%	38.15%	35.57%	36.82%	41.40%	41.17%	41.54%	38.15%	35.57%

T403 Query 7 idf

T403 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.57%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%	90%
20%	0.41%	78.26%	78.26%	78.26%	78.26%	69.23%	78.26%	78.26%	78.26%	75%	78.26%	69.23%	81.82%	78.26%	78.26%	75%	78.26%	69.23%
30%	0.44%	67.50%	75%	61.36%	60%	60%	71.05%	75%	75%	64.29%	60%	58.70%	55.10%	60%	57.45%	64.29%	60%	58.70%
40%	0.40%	63.16%	65.46%	43.90%	36.74%	43.90%	60%	64.29%	64.29%	50.70%	35.64%	43.37%	41.38%	43.90%	45%	50.70%	35.64%	43.37%
50%	0.38%	52.87%	42.59%	40.71%	31.51%	38.98%	40.35%	46%	45.55%	44.23%	30.67%	36.80%	42.99%	43.81%	40.71%	44.23%	30.67%	36.80%
60%	0.45%	36.18%	34.38%	29.57%	14.83%	24.34%	39.86%	42.64%	42.31%	30.90%	23.91%	27.50%	33.13%	32.74%	35.48%	30.90%	23.91%	27.36%
70%	0.33%	25.70%	24.52%	19.34%	8.18%	17.20%	31.37%	35.36%	34.41%	19.88%	11.81%	17.39%	20.92%	22.78%	19.57%	19.81%	11.77%	17.34%
80%	0.29%	15.50%	15.34%	13.70%	4.24%	10.27%	21.22%	21.04%	20.51%	14.46%	5.75%	14.51%	14.84%	16.55%	15.53%	14.43%	5.74%	14.46%
90%	0%	0%	0%	0%	0%	0%	14.64%	16.91%	15.86%	11.97%	4.18%	9.56%	11.85%	13.88%	11.47%	11.90%	4.17%	9.51%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1.09%	0.31%	0.32%	1.19%	1.81%	0.40%
Avg30	0.48%	78.59%	81.09%	76.54%	76.09%	73.08%	79.77%	81.09%	81.09%	76.43%	76.09%	72.64%	75.64%	76.09%	75.24%	76.43%	76.09%	72.64%
AVG100	0.33%	42.92%	42.55%	37.68%	32.38%	35.39%	44.68%	46.95%	46.62%	40.14%	34.02%	36.71%	39.31%	40.22%	39.38%	40.25%	34.20%	36.72%

T403 Query 8 idf

T403 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.54%	81.82%	81.82%	81.82%	52.94%	64.29%	75%	81.82%	81.82%	81.82%	52.94%	64.29%	81.82%	75%	75%	81.82%	52.94%	64.29%
20%	0.75%	85.71%	85.71%	75%	50%	33.33%	81.82%	85.71%	85.71%	72%	50%	32.73%	75%	78.26%	78.26%	72%	50%	32.73%
30%	0.88%	84.38%	77.14%	60%	32.14%	30.68%	81.82%	77.14%	77.14%	67.50%	32.14%	29.35%	57.45%	64.29%	64.29%	67.50%	32.14%	29.35%
40%	0.75%	61.02%	67.93%	46.75%	26.67%	26.67%	63.16%	67.93%	67.93%	61.02%	26.47%	24.66%	58.07%	66.67%	64.29%	61.02%	26.47%	24.66%
50%	0.85%	43.40%	44.66%	33.09%	16.43%	23.71%	38.66%	45.10%	45.10%	44.23%	15.81%	21.91%	43.81%	42.59%	40.71%	44.23%	15.81%	21.80%
60%	0.61%	14.25%	29.41%	11.63%	17.30%	21.91%	32.35%	41.99%	41.99%	31.79%	16.98%	20%	29.89%	30.73%	26.96%	31.61%	16.98%	19.93%
70%	0%	8.11%	13.17%	10.18%	11.62%	17.44%	31.07%	32.99%	31.68%	12.72%	14.10%	18.71%	14.07%	14.04%	13.25%	12.65%	14.07%	18.61%
80%	0%	0%	0%	0%	1.24%	7.28%	9.46%	23.93%	22.12%	12.72%	3.04%	14.93%	12.11%	13.67%	12.48%	12.61%	3.03%	14.78%
90%	0%	0%	0%	0%	0%	0%	6.99%	15.07%	13.00%	10.68%	1.37%	7.95%	10.35%	8.88%	8.24%	10.39%	1.36%	7.82%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.06%	83.97%	81.56%	72.27%	45.03%	42.77%	79.55%	81.56%	81.56%	73.77%	45.03%	42.12%	71.42%	72.52%	72.52%	73.77%	45.03%	42.12%
AVG100	0.54%	37.87%	39.98%	31.85%	20.83%	22.53%	42.03%	47.17%	46.65%	39.45%	21.28%	23.45%	38.26%	39.41%	38.35%	39.38%	21.28%	23.40%

T403 Query 9 idf

T403 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.99%	60%	81.82%	81.82%	81.82%	81.82%	75%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	81.82%	60%	81.82%
20%	0.96%	64.29%	78.26%	90%	40.91%	62.07%	72%	81.82%	85.71%	85.71%	41.86%	69.23%	75%	81.82%	81.82%	81.82%	60%	66.67%
30%	0.99%	51.92%	61.36%	56.25%	31.03%	40.91%	64.29%	71.05%	62.79%	56.25%	32.14%	49.09%	61.36%	62.79%	55.10%	62.79%	24.55%	44.26%
40%	0.92%	40%	41.38%	49.32%	25.71%	32.43%	52.94%	67.93%	64.29%	51.43%	27.69%	40.45%	47.37%	50%	46.75%	52.94%	25.71%	38.30%
50%	0.81%	31.72%	25.28%	38.98%	23.12%	30.07%	46.94%	58.23%	48.42%	45.10%	24.47%	36.51%	41.44%	43.81%	40.71%	44.23%	25.70%	35.66%
60%	0%	15.49%	16.32%	26.57%	16.52%	21.57%	41.05%	42.64%	44%	40.44%	23.01%	35.71%	35.48%	37.16%	35.48%	36.18%	19.30%	34.16%
70%	0%	0%	8.85%	13.01%	6.10%	11.19%	30.92%	31.37%	31.68%	28.96%	19.94%	24.62%	29.09%	28.07%	27.00%	26.67%	18.71%	25.10%
80%	0%	0%	0%	0%	0%	0%	19.26%	21.10%	19.73%	17.94%	15.53%	18.81%	15.84%	18.96%	16.12%	16.67%	14.37%	16.78%
90%	0%	0%	0%	0%	0%	0%	16.60%	17.52%	16.80%	13.16%	6.23%	15.47%	12.33%	14.19%	12.50%	12.81%	8.76%	15.19%
100%	0%	0%	0%	0%	0%	0%	3.83%	1.07%	3.73%	3.98%	1.39%	1.18%	3.71%	1.06%	4.23%	3.69%	0.49%	1.16%
Avg30	0.98%	58.74%	73.81%	76.02%	51.25%	61.60%	70.43%	78.23%	76.77%	74.59%	51.94%	66.71%	72.73%	75.48%	72.91%	75.48%	48.18%	64.25%
AVG100	0.47%	26.34%	31.33%	35.59%	22.52%	28.01%	42.28%	47.45%	45.90%	42.48%	27.41%	37.29%	40.34%	41.97%	40.15%	41.96%	25.76%	35.91%

T403 Query 10 idf

T403 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	90%	100%	81.82%	90%	90%	90%	81.82%	81.82%	81.82%	90%	90%	90%	90%	90%	90%	90%	90%	90%
20%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%	85.71%
30%	72.97%	57.45%	90%	84.38%	84.38%	84.38%	84.38%	90%	90%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%	84.38%
40%	45.57%	38.30%	76.60%	78.26%	80%	78.26%	81.82%	83.72%	76.60%	78.26%	80%	78.26%	80%	78.26%	78.26%	78.26%	80%	78.26%
50%	11.33%	7.65%	14.60%	66.67%	69.70%	69.70%	76.67%	15.13%	15.23%	63.89%	66.67%	69.70%	63.89%	69.70%	63.89%	63.89%	66.67%	69.70%
60%	4.68%	4.13%	9.91%	46.22%	50%	44%	40.74%	16.13%	10%	49.55%	51.40%	42.97%	49.55%	42.97%	49.55%	49.55%	51.40%	42.97%
70%	0.96%	1.00%	9.71%	28.19%	21.41%	17.07%	16.20%	8.89%	9.79%	35.36%	34.04%	28.83%	32.32%	28.83%	34.60%	35.36%	34.04%	28.83%
80%	0%	0%	8.37%	9.01%	7.47%	5.16%	14.12%	0.35%	9.32%	24.83%	20.22%	9.44%	23.40%	9.31%	23.86%	24.83%	20.51%	9.44%
90%	0%	0%	0%	0%	0%	0%	9.40%	0.39%	9.01%	10.39%	9.76%	5.24%	10.53%	5.27%	9.54%	10.39%	9.65%	5.24%
100%	0%	0%	0%	0%	0%	0%	4.14%	0.43%	3.88%	1.59%	1.58%	0.67%	1.46%	0.72%	1.59%	1.49%	1.45%	0.67%
Avg30	82.90%	81.05%	85.84%	86.70%	86.70%	86.70%	83.97%	85.84%	85.84%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%	86.70%
AVG100	31.12%	29.42%	37.67%	48.84%	48.87%	47.43%	49.50%	38.26%	39.14%	52.40%	52.38%	49.52%	52.12%	49.51%	52.14%	52.39%	52.38%	49.52%

T416 Query 1

T416 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	20%	25%	23.08%	42.86%	12%	50%	17.65%	23.08%	23.08%	42.86%	12%	50%	21.43%	25%	25%	42.86%	12%	50%
30%	25%	30.77%	28.57%	50%	15.39%	57.14%	22.22%	28.57%	28.57%	50%	15.39%	57.14%	26.67%	30.77%	30.77%	50%	15.39%	57.14%
40%	33.33%	40%	37.50%	60%	21.43%	66.67%	30%	37.50%	37.50%	60%	21.43%	66.67%	35.29%	40%	40%	60%	21.43%	66.67%
50%	36.84%	43.75%	41.18%	63.64%	24.14%	70%	33.33%	41.18%	41.18%	63.64%	24.14%	70%	38.89%	43.75%	43.75%	63.64%	24.14%	70%
60%	40%	47.06%	44.44%	66.67%	26.67%	72.73%	36.36%	44.44%	44.44%	66.67%	26.67%	72.73%	42.11%	47.06%	47.06%	66.67%	26.67%	72.73%
70%	22.73%	37.04%	45.46%	47.62%	2.04%	34.48%	22.73%	45.46%	45.46%	47.62%	2.04%	34.48%	30.30%	41.67%	41.67%	47.62%	1.19%	34.48%
80%	23.40%	28.95%	32.35%	33.33%	2.24%	36.67%	23.40%	32.35%	32.35%	33.33%	2.24%	36.67%	30.56%	33.33%	33.33%	33.33%	1.31%	36.67%
90%	17.81%	13.83%	15.12%	15.85%	1.67%	0.20%	17.57%	15.12%	15.12%	15.85%	1.67%	0.20%	20.64%	13.68%	13.68%	15.66%	1.15%	0.20%
100%	0.20%	0.22%	0.22%	0.20%	0.16%	0.12%	0.20%	0.22%	0.22%	0.20%	0.16%	0.12%	0.17%	0.21%	0.21%	0.19%	0.15%	0.11%
Avg30	48.33%	51.92%	50.55%	64.29%	42.46%	69.05%	46.62%	50.55%	50.55%	64.29%	42.46%	69.05%	49.37%	51.92%	51.92%	64.29%	42.46%	69.05%
AVG100	31.93%	36.66%	36.79%	48.02%	20.57%	48.80%	30.35%	36.79%	36.79%	48.02%	20.57%	48.80%	34.61%	37.55%	37.55%	48.00%	20.34%	48.80%

T416 Query 2

T416 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	25%	75%	75%	75%	100%	75%	75%	75%	75%	75%	100%	75%	75%	75%	75%	75%	100%	75%
30%	30.77%	80%	80%	80%	100%	80%	80%	80%	80%	80%	100%	80%	80%	80%	80%	80%	100%	80%
40%	35.29%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%
50%	38.89%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%
60%	42.11%	88.89%	88.89%	88.89%	28.57%	88.89%	88.89%	88.89%	88.89%	88.89%	28.57%	88.89%	88.89%	88.89%	88.89%	88.89%	28.57%	88.89%
70%	47.62%	76.92%	66.67%	66.67%	4.24%	66.67%	76.92%	66.67%	66.67%	66.67%	4.15%	66.67%	76.92%	66.67%	66.67%	66.67%	4.15%	66.67%
80%	50%	50%	19.64%	19.64%	1.72%	23.91%	78.57%	17.19%	18.97%	18.97%	1.71%	23.91%	78.57%	17.19%	18.97%	18.97%	1.71%	23.91%
90%	16.67%	1.22%	0.47%	0.47%	1.31%	8.03%	50%	7.43%	8.44%	8.44%	1.30%	7.69%	50%	7.43%	8.44%	8.44%	1.30%	7.69%
100%	0%	0%	0%	0%	0.52%	0.55%	37.84%	0.50%	0.48%	0.48%	0.49%	0.55%	37.84%	0.50%	0.48%	0.48%	0.49%	0.55%
Avg30	51.92%	85.00%	85.00%	85.00%	100.00%	85.00%	85.00%	85.00%	85.00%	85.00%	100.00%	85.00%	85.00%	85.00%	85.00%	85.00%	100.00%	85.00%
AVG100	38.63%	64.52%	60.39%	60.39%	53.64%	61.63%	76.04%	60.89%	61.17%	61.17%	53.62%	61.59%	76.04%	60.89%	61.17%	61.17%	53.62%	61.59%

T416 Query 3

T416 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.98%	0.98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	1.45%	1.45%	100%	100%	100%	50%	100%	100%	100%	100%	100%	50%	100%	100%	100%	100%	100%	50%
30%	1.79%	1.79%	100%	100%	100%	57.14%	100%	100%	100%	100%	100%	57.14%	100%	100%	100%	100%	100%	57.14%
40%	0.67%	0.67%	100%	100%	100%	66.67%	100%	100%	100%	100%	100%	66.67%	100%	100%	100%	100%	100%	66.67%
50%	0.72%	0.72%	100%	100%	100%	70%	100%	100%	100%	100%	100%	70%	100%	100%	100%	100%	100%	70%
60%	0.79%	0.79%	80%	80%	100%	72.73%	66.67%	61.54%	80%	80%	100%	72.73%	66.67%	61.54%	80%	80%	100%	72.73%
70%	0.90%	0.90%	62.50%	62.50%	76.92%	55.56%	71.43%	66.67%	62.50%	62.50%	76.92%	55.56%	71.43%	66.67%	62.50%	62.50%	76.92%	55.56%
80%	0.93%	0.93%	55%	55%	68.75%	57.90%	45.83%	55%	55%	55%	68.75%	57.90%	45.83%	55%	55%	55%	68.75%	57.90%
90%	0.89%	0.89%	40.63%	40.63%	26.53%	46.43%	50%	39.39%	40.63%	40.63%	26.53%	46.43%	50%	39.39%	40.63%	40.63%	26.53%	46.43%
100%	0.87%	0.87%	1.00%	1.00%	1.70%	20.59%	1.16%	1.57%	0.69%	0.69%	1.70%	20.59%	1.16%	1.57%	0.69%	0.69%	1.70%	20.59%
Avg30	1.41%	1.41%	100.00%	100.00%	100.00%	69.05%	100.00%	100.00%	100.00%	100.00%	100.00%	69.05%	100.00%	100.00%	100.00%	100.00%	100.00%	69.05%
AVG100	1.00%	1.00%	73.91%	73.91%	77.39%	59.70%	73.51%	72.42%	73.88%	73.88%	77.39%	59.70%	73.51%	72.42%	73.88%	73.88%	77.39%	59.70%

T416 Query 4

T416 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	50%	75%	75%	75%	10%	100%	50%	75%	75%	75%	10%	100%	50%	75%	75%	75%	10%	100%
30%	57.14%	80%	80%	80%	12.90%	100%	57.14%	80%	80%	80%	12.90%	100%	57.14%	80%	80%	80%	12.90%	100%
40%	66.67%	85.71%	85.71%	85.71%	18.18%	100%	66.67%	85.71%	85.71%	85.71%	18.18%	100%	66.67%	85.71%	85.71%	85.71%	18.18%	100%
50%	70%	87.50%	87.50%	87.50%	20.59%	100%	70%	87.50%	87.50%	87.50%	20.59%	100%	70%	87.50%	87.50%	87.50%	20.59%	100%
60%	72.73%	88.89%	88.89%	88.89%	22.86%	100%	72.73%	88.89%	88.89%	88.89%	22.86%	100%	72.73%	88.89%	88.89%	88.89%	22.86%	100%
70%	11.49%	21.74%	21.74%	21.74%	0.78%	0.33%	11.49%	21.74%	21.74%	21.74%	0.78%	0.33%	11.49%	21.74%	21.74%	21.74%	0.78%	0.33%
80%	0.39%	0.39%	0.39%	0.39%	0.38%	0.33%	0.39%	0.39%	0.39%	0.39%	0.38%	0.33%	0.39%	0.39%	0.39%	0.39%	0.38%	0.33%
90%	0.40%	0.40%	0.40%	0.40%	0.05%	0.05%	0.40%	0.40%	0.40%	0.40%	0.05%	0.05%	0.40%	0.40%	0.40%	0.40%	0.05%	0.05%
100%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%
Avg30	69.05%	85.00%	85.00%	85.00%	40.97%	100.00%	69.05%	85.00%	85.00%	85.00%	40.97%	100.00%	69.05%	85.00%	85.00%	85.00%	40.97%	100.00%
AVG100	42.88%	53.96%	53.96%	53.96%	18.58%	60.08%	42.88%	53.96%	53.96%	53.96%	18.58%	60.08%	42.88%	53.96%	53.96%	53.96%	18.58%	60.08%

T416 Query 5

T416 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.16%	0.16%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.36%	0.36%	75%	75%	42.86%	100%	50%	75%	50%	50%	42.86%	100%	50%	75%	50%	50%	42.86%	100%
30%	0.46%	0.46%	66.67%	66.67%	50%	100%	57.14%	80%	57.14%	57.14%	50%	100%	57.14%	80%	57.14%	57.14%	50%	100%
40%	0.29%	0.29%	75%	75%	60%	0.20%	40%	85.71%	66.67%	66.67%	60%	0.20%	40%	85.71%	66.67%	66.67%	60%	0.20%
50%	0.34%	0.34%	77.78%	77.78%	63.64%	0.24%	43.75%	87.50%	70%	70%	63.64%	0.24%	43.75%	87.50%	70%	70%	63.64%	0.24%
60%	0.38%	0.38%	80%	80%	66.67%	0.27%	47.06%	88.89%	72.73%	72.73%	66.67%	0.27%	47.06%	88.89%	72.73%	72.73%	66.67%	0.27%
70%	0.42%	0.42%	83.33%	83.33%	20.83%	0.34%	52.63%	41.67%	50%	50%	20.83%	0.34%	52.63%	41.67%	50%	50%	20.83%	0.34%
80%	0.43%	0.43%	73.33%	73.33%	22.45%	0.37%	55%	44%	28.21%	28.21%	22.45%	0.37%	55%	44%	28.21%	28.21%	22.45%	0.37%
90%	0.45%	0.45%	22.81%	22.81%	0.75%	0.34%	8.61%	0.48%	1.16%	1.16%	0.73%	0.27%	8.61%	0.48%	1.16%	1.16%	0.73%	0.27%
100%	0%	0%	1.41%	1.41%	0.04%	0.36%	1.13%	0.42%	0.33%	0.33%	0.04%	0.29%	1.13%	0.42%	0.33%	0.33%	0.04%	0.29%
Avg30	0.33%	0.33%	80.56%	80.56%	64.29%	100.00%	69.05%	85.00%	69.05%	69.05%	64.29%	100.00%	69.05%	85.00%	69.05%	69.05%	64.29%	100.00%
AVG100	0.33%	0.33%	65.53%	65.53%	42.72%	30.21%	45.53%	60.37%	49.62%	49.62%	42.72%	30.20%	45.53%	60.37%	49.62%	49.62%	42.72%	30.20%

T416 Query 6

T416 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	50%	50%	50%	50%	50%	100%	50%	50%	50%	50%	50%	100%	50%	50%	50%	50%	50%	100%
20%	9.38%	12%	10.35%	10.35%	0.34%	0.04%	8.11%	6.98%	10.35%	10.35%	0.30%	0.04%	8.11%	6.98%	10.35%	10.35%	0.30%	0.04%
30%	12.12%	15.39%	13.33%	13.33%	0.45%	0.03%	10.53%	9.09%	13.33%	13.33%	0.39%	0.03%	10.53%	9.09%	13.33%	13.33%	0.39%	0.03%
40%	17.14%	21.43%	18.75%	18.75%	0.67%	0.05%	15%	13.04%	18.75%	18.75%	0.59%	0.05%	15%	13.04%	18.75%	18.75%	0.59%	0.05%
50%	19.44%	24.14%	21.21%	21.21%	0.78%	0.06%	17.07%	14.89%	21.21%	21.21%	0.69%	0.06%	17.07%	14.89%	21.21%	21.21%	0.69%	0.06%
60%	21.62%	26.67%	23.53%	23.53%	0.89%	0.06%	19.05%	16.67%	23.53%	23.53%	0.79%	0.06%	19.05%	16.67%	23.53%	23.53%	0.79%	0.06%
70%	0.18%	0.13%	0.13%	0.13%	0.11%	0.08%	0.17%	0.13%	0.13%	0.13%	0.11%	0.08%	0.17%	0.13%	0.13%	0.13%	0.11%	0.08%
80%	0.13%	0.09%	0.09%	0.09%	0.03%	0.09%	0.13%	0.09%	0.09%	0.09%	0.03%	0.09%	0.13%	0.09%	0.09%	0.09%	0.03%	0.09%
90%	0.14%	0.10%	0.10%	0.10%	0.03%	0.10%	0.12%	0.10%	0.10%	0.10%	0.03%	0.10%	0.12%	0.10%	0.10%	0.10%	0.03%	0.10%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	23.83%	25.80%	24.56%	24.56%	16.93%	33.36%	22.88%	22.02%	24.56%	24.56%	16.90%	33.36%	22.88%	22.02%	24.56%	24.56%	16.90%	33.36%
AVG100	13.02%	14.99%	13.75%	13.75%	5.33%	10.05%	12.02%	11.10%	13.75%	13.75%	5.29%	10.05%	12.02%	11.10%	13.75%	13.75%	5.29%	10.05%

T416 Query 7

T416 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.05%	0.05%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30%	0.06%	0.06%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
40%	0.01%	0.01%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
50%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
60%	0.02%	0.02%	16.33%	16.33%	22.86%	7.41%	17.02%	17.78%	16.33%	16.33%	22.86%	7.34%	17.02%	17.78%	16.33%	16.33%	22.86%	7.34%
70%	0.02%	0.02%	0.02%	0.02%	0.04%	0.04%	0.03%	0.02%	0.03%	0.03%	0.05%	0.14%	0.03%	0.02%	0.03%	0.03%	0.05%	0.14%
80%	0.02%	0.02%	0.02%	0.02%	0.02%	0.05%	0.03%	0.02%	0.03%	0.03%	0.04%	0.05%	0.03%	0.02%	0.03%	0.03%	0.04%	0.05%
90%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.03%	0.03%	0.02%	0.02%	0.02%	0.02%	0.03%	0.03%
100%	0%	0%	0%	0%	0%	0%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%
Avg30	0.04%	0.04%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AVG100	0.02%	0.02%	51.64%	51.64%	52.29%	50.75%	51.71%	51.79%	51.64%	51.64%	52.30%	50.76%	51.71%	51.79%	51.64%	51.64%	52.30%	50.76%

T416 Query 8

T416 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	15.79%	16.67%	33.33%	33.33%	9.38%	1.22%	33.33%	33.33%	33.33%	33.33%	9.09%	1.22%	33.33%	30%	30%	33.33%	9.09%	1.22%
30%	20%	21.05%	40%	40%	12.12%	1.62%	40%	40%	40%	40%	11.77%	1.61%	40%	36.36%	36.36%	40%	11.77%	1.61%
40%	27.27%	28.57%	50%	50%	17.14%	2.41%	50%	50%	50%	50%	16.67%	2.40%	50%	46.15%	46.15%	50%	16.67%	2.40%
50%	30.44%	31.82%	53.85%	53.85%	19.44%	2.80%	53.85%	53.85%	53.85%	53.85%	18.92%	2.79%	53.85%	50%	50%	53.85%	18.92%	2.79%
60%	33.33%	34.78%	57.14%	57.14%	21.62%	3.19%	57.14%	57.14%	57.14%	57.14%	21.05%	3.18%	57.14%	53.33%	53.33%	57.14%	21.05%	3.18%
70%	1.37%	0.05%	0.10%	0.09%	0.06%	0.15%	0.10%	0.10%	0.10%	0.09%	0.06%	0.15%	0.09%	0.09%	0.09%	0.09%	0.06%	0.14%
80%	0.47%	0.06%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
90%	0.06%	0.03%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	45.26%	45.91%	57.78%	57.78%	40.50%	34.28%	57.78%	57.78%	57.78%	57.78%	40.29%	34.28%	57.78%	55.45%	55.45%	57.78%	40.29%	34.28%
AVG100	22.87%	23.30%	33.45%	33.45%	17.98%	11.14%	33.45%	33.45%	33.45%	33.45%	17.76%	11.14%	33.45%	31.60%	31.60%	33.45%	17.76%	11.14%

T416 Query 9

T416 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.23%	0.23%	100%	100%	9.09%	33.33%	100%	100%	100%	100%	9.09%	33.33%	100%	100%	100%	100%	9.09%	33.33%
20%	0.29%	0.29%	100%	100%	23.08%	60%	37.50%	100%	100%	100%	23.08%	60%	37.50%	100%	100%	100%	23.08%	60%
30%	0.23%	0.23%	100%	100%	28.57%	66.67%	44.44%	100%	100%	100%	28.57%	66.67%	44.44%	100%	100%	100%	28.57%	66.67%
40%	0.33%	0.33%	100%	100%	37.50%	75%	54.55%	100%	100%	100%	37.50%	75%	54.55%	100%	100%	100%	37.50%	75%
50%	0.36%	0.36%	100%	100%	41.18%	77.78%	58.33%	100%	100%	100%	41.18%	77.78%	58.33%	100%	100%	100%	41.18%	77.78%
60%	0.39%	0.39%	66.67%	66.67%	0.97%	2.85%	61.54%	100%	66.67%	66.67%	0.72%	1.63%	61.54%	100%	66.67%	66.67%	0.72%	1.63%
70%	0%	0%	0%	0%	0.50%	1.83%	0.54%	0.28%	0.56%	0.56%	0.39%	1.00%	0.54%	0.28%	0.56%	0.56%	0.39%	1.00%
80%	0%	0%	0%	0%	0.06%	0.06%	0.23%	0.23%	0.23%	0.23%	0.06%	0.28%	0.23%	0.23%	0.23%	0.23%	0.06%	0.28%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.04%	0.07%	0%	0%	0%	0%	0.04%	0.07%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%
Avg30	0.25%	0.25%	100.00%	100.00%	20.25%	53.33%	60.65%	100.00%	100.00%	100.00%	20.25%	53.33%	60.65%	100.00%	100.00%	100.00%	20.25%	53.33%
AVG100	0.18%	0.18%	56.67%	56.67%	14.09%	31.75%	35.71%	60.05%	56.75%	56.75%	14.07%	31.58%	35.71%	60.05%	56.75%	56.75%	14.07%	31.58%

T416 Query 10

T416 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	10.35%	20%	20%	21.43%	0.17%	0.10%	10.35%	20%	20%	21.43%	0.17%	0.10%	12%	15%	15%	21.43%	0.17%	0.10%
30%	13.33%	25%	25%	26.67%	0.22%	0.14%	13.33%	25%	25%	26.67%	0.22%	0.14%	15.39%	19.05%	19.05%	26.67%	0.22%	0.14%
40%	18.75%	33.33%	33.33%	35.29%	0.33%	0.20%	18.75%	33.33%	33.33%	35.29%	0.33%	0.20%	21.43%	26.09%	26.09%	35.29%	0.33%	0.20%
50%	21.21%	36.84%	36.84%	38.89%	0.39%	0.24%	21.21%	36.84%	36.84%	38.89%	0.39%	0.24%	24.14%	29.17%	29.17%	38.89%	0.39%	0.24%
60%	23.53%	40%	40%	42.11%	0.44%	0.27%	23.53%	40%	40%	42.11%	0.44%	0.27%	26.67%	32%	32%	42.11%	0.44%	0.27%
70%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.05%	0.05%
80%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.04%	0.04%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	41.23%	48.33%	48.33%	49.37%	33.46%	33.41%	41.23%	48.33%	48.33%	49.37%	33.46%	33.41%	42.46%	44.68%	44.68%	49.37%	33.46%	33.41%
AVG100	18.72%	25.52%	25.52%	26.44%	10.16%	10.10%	18.72%	25.52%	25.52%	26.44%	10.16%	10.10%	19.96%	22.13%	22.13%	26.44%	10.16%	10.10%

T416 Query 1 entropy

T416 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	37.50%	60%	60%	75%	30%	60%	37.50%	60%	60%	75%	30%	60%	50%	75%	75%	75%	30%	60%
30%	44.44%	66.67%	66.67%	80%	36.36%	66.67%	44.44%	66.67%	66.67%	80%	36.36%	66.67%	57.14%	80%	80%	80%	36.36%	66.67%
40%	54.55%	75%	75%	85.71%	46.15%	75%	54.55%	75%	75%	85.71%	46.15%	75%	66.67%	85.71%	85.71%	85.71%	46.15%	75%
50%	58.33%	77.78%	77.78%	87.50%	50%	77.78%	58.33%	77.78%	77.78%	87.50%	50%	77.78%	70%	87.50%	87.50%	87.50%	50%	77.78%
60%	61.54%	80%	80%	88.89%	53.33%	80%	61.54%	80%	80%	88.89%	53.33%	80%	72.73%	88.89%	88.89%	88.89%	53.33%	80%
70%	32.26%	52.63%	58.82%	71.43%	1.10%	2.32%	34.48%	58.82%	58.82%	71.43%	1.10%	2.32%	35.71%	71.43%	71.43%	71.43%	0.81%	2.32%
80%	28.95%	32.35%	44%	45.83%	1.14%	2.54%	29.73%	44%	44%	45.83%	1.14%	2.54%	33.33%	42.31%	42.31%	45.83%	0.84%	2.54%
90%	16.46%	15.66%	18.57%	20.97%	0.21%	0.47%	16.25%	18.57%	18.57%	20.97%	0.21%	0.47%	19.12%	18.31%	18.31%	20.97%	0.20%	0.47%
100%	0.22%	0.26%	0.25%	0.24%	0.19%	0.29%	0.22%	0.25%	0.25%	0.24%	0.19%	0.29%	0.19%	0.25%	0.25%	0.21%	0.18%	0.29%
Avg30	60.65%	75.56%	75.56%	85.00%	55.45%	75.56%	60.65%	75.56%	75.56%	85.00%	55.45%	75.56%	69.05%	85.00%	85.00%	85.00%	55.45%	75.56%
AVG100	43.42%	56.04%	58.11%	65.56%	31.85%	46.51%	43.70%	58.11%	58.11%	65.56%	31.85%	46.51%	50.49%	64.94%	64.94%	65.55%	31.79%	46.51%

T416 Query 2 entropy

T416 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	33.33%	75%	75%	75%	75%	75%	75%	75%	75%	75%	100%	75%	75%	75%	75%	100%	100%	75%
30%	30.77%	80%	80%	80%	80%	80%	80%	80%	80%	80%	100%	80%	80%	80%	80%	100%	100%	80%
40%	40%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%
50%	43.75%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%
60%	47.06%	88.89%	88.89%	88.89%	7.21%	88.89%	88.89%	88.89%	88.89%	88.89%	7.08%	88.89%	88.89%	88.89%	88.89%	88.89%	7.08%	88.89%
70%	52.63%	66.67%	66.67%	66.67%	3.82%	66.67%	76.92%	66.67%	66.67%	66.67%	3.75%	66.67%	76.92%	66.67%	66.67%	66.67%	3.75%	66.67%
80%	55%	20.37%	18.33%	18.33%	1.54%	20.76%	78.57%	15.49%	16.42%	16.42%	1.53%	20.76%	78.57%	15.49%	16.42%	16.42%	1.53%	20.76%
90%	24.53%	0.49%	0.49%	0.49%	1.39%	9.22%	50%	8.39%	9.56%	9.56%	1.38%	8.73%	50%	8.39%	9.56%	9.56%	1.38%	8.73%
100%	0%	0%	0%	0%	0.51%	0.57%	43.75%	0.52%	0.49%	0.49%	0.49%	0.56%	43.75%	0.52%	0.49%	0.49%	0.49%	0.56%
Avg30	54.70%	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	85.00%	100.00%	85.00%	85.00%	85.00%	85.00%	100.00%	100.00%	85.00%
AVG100	42.71%	60.46%	60.26%	60.26%	46.95%	61.43%	76.63%	60.82%	61.02%	61.02%	51.42%	61.38%	76.63%	60.82%	61.02%	65.52%	51.42%	61.38%

T416 Query 3 entropy

T416 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	0.98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	75%	1.45%	100%	100%	75%	75%	100%	100%	100%	100%	75%	75%	100%	100%	100%	100%	75%	75%
30%	1.79%	1.79%	100%	100%	80%	80%	100%	100%	100%	100%	80%	80%	100%	100%	100%	100%	80%	80%
40%	0.67%	0.67%	100%	100%	85.71%	85.71%	100%	100%	100%	100%	85.71%	85.71%	100%	100%	100%	100%	85.71%	85.71%
50%	0.72%	0.72%	100%	100%	87.50%	87.50%	100%	100%	100%	100%	87.50%	87.50%	100%	100%	100%	100%	87.50%	87.50%
60%	0.79%	0.79%	88.89%	88.89%	88.89%	88.89%	66.67%	61.54%	88.89%	88.89%	88.89%	88.89%	66.67%	61.54%	88.89%	88.89%	88.89%	88.89%
70%	0.90%	0.90%	62.50%	62.50%	58.82%	55.56%	66.67%	66.67%	62.50%	62.50%	58.82%	55.56%	66.67%	66.67%	62.50%	62.50%	58.82%	55.56%
80%	0.93%	0.93%	55%	55%	61.11%	57.90%	45.83%	57.90%	55%	55%	61.11%	57.90%	45.83%	57.90%	55%	55%	61.11%	57.90%
90%	0.89%	0.89%	40.63%	40.63%	22.41%	44.83%	50%	39.39%	40.63%	40.63%	22.41%	44.83%	50%	39.39%	40.63%	40.63%	22.41%	44.83%
100%	0.87%	0.87%	1.02%	1.02%	1.70%	20.90%	1.31%	1.59%	0.69%	0.69%	1.70%	20.90%	1.31%	1.59%	0.69%	0.69%	1.70%	20.90%
Avg30	58.93%	1.41%	100.00%	100.00%	85.00%	85.00%	100.00%	100.00%	100.00%	100.00%	85.00%	85.00%	100.00%	100.00%	100.00%	100.00%	85.00%	85.00%
AVG100	18.26%	1.00%	74.80%	74.80%	66.12%	69.63%	73.05%	72.71%	74.77%	74.77%	66.12%	69.63%	73.05%	72.71%	74.77%	74.77%	66.12%	69.63%

T416 Query 4 entropy

T416 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	60%	100%	100%	100%	10.71%	100%	60%	100%	100%	100%	10.71%	100%	60%	100%	100%	100%	10.71%	100%
30%	66.67%	100%	100%	100%	13.79%	100%	66.67%	100%	100%	100%	13.79%	100%	66.67%	100%	100%	100%	13.79%	100%
40%	75%	100%	100%	100%	19.36%	100%	75%	100%	100%	100%	19.36%	100%	75%	100%	100%	100%	19.36%	100%
50%	77.78%	100%	100%	100%	21.88%	100%	77.78%	100%	100%	100%	21.88%	100%	77.78%	100%	100%	100%	21.88%	100%
60%	80%	100%	100%	100%	24.24%	100%	80%	100%	100%	100%	24.24%	100%	80%	100%	100%	100%	24.24%	100%
70%	13.16%	30.30%	30.30%	30.30%	0.93%	0.33%	13.16%	30.30%	30.30%	30.30%	0.93%	0.33%	13.16%	30.30%	30.30%	30.30%	0.93%	0.33%
80%	0.40%	0.40%	0.40%	0.40%	0.34%	0.33%	0.40%	0.40%	0.40%	0.40%	0.34%	0.33%	0.40%	0.40%	0.40%	0.40%	0.34%	0.33%
90%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%
100%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%
Avg30	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%
AVG100	47.34%	63.11%	63.11%	63.11%	19.13%	60.08%	47.34%	63.11%	63.11%	63.11%	19.13%	60.08%	47.34%	63.11%	63.11%	63.11%	19.13%	60.08%

T416 Query 5 entropy

T416 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.16%	0.16%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.36%	0.36%	75%	75%	42.86%	100%	60%	100%	60%	60%	42.86%	100%	60%	100%	100%	100%	42.86%	100%
30%	0.46%	0.46%	66.67%	66.67%	50%	100%	66.67%	100%	66.67%	66.67%	50%	100%	66.67%	100%	100%	100%	50%	100%
40%	0.29%	0.29%	75%	75%	60%	0.20%	42.86%	100%	75%	75%	60%	0.20%	42.86%	100%	75%	75%	60%	0.20%
50%	0.34%	0.34%	77.78%	77.78%	63.64%	0.24%	46.67%	100%	77.78%	77.78%	63.64%	0.24%	46.67%	100%	77.78%	77.78%	63.64%	0.24%
60%	0.38%	0.38%	80%	80%	66.67%	0.27%	50%	100%	80%	80%	66.67%	0.27%	50%	100%	80%	80%	66.67%	0.27%
70%	0.42%	0.42%	83.33%	83.33%	17.24%	0.34%	55.56%	50%	58.82%	58.82%	17.24%	0.34%	55.56%	50%	58.82%	58.82%	17.24%	0.34%
80%	0.43%	0.43%	73.33%	73.33%	18.64%	0.37%	57.90%	21.57%	17.74%	17.74%	18.64%	0.37%	57.90%	21.57%	17.74%	17.74%	18.64%	0.37%
90%	0.45%	0.45%	27.66%	27.66%	0.77%	0.34%	3.99%	0.46%	1.33%	1.33%	0.75%	0.27%	3.99%	0.46%	1.33%	1.33%	0.75%	0.27%
100%	0%	0%	1.86%	1.86%	0.04%	0.36%	1.34%	0.40%	0.31%	0.31%	0.04%	0.29%	1.34%	0.40%	0.31%	0.31%	0.04%	0.29%
Avg30	0.33%	0.33%	80.56%	80.56%	64.29%	100.00%	75.56%	100.00%	75.56%	75.56%	64.29%	100.00%	75.56%	100.00%	100.00%	100.00%	64.29%	100.00%
AVG100	0.33%	0.33%	66.06%	66.06%	41.99%	30.21%	48.50%	67.24%	53.77%	53.77%	41.98%	30.20%	48.50%	67.24%	61.10%	61.10%	41.98%	30.20%

T416 Query 6 entropy

T416 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	50%	100%	100%	100%	100%	100%	50%	100%	100%	100%	100%
20%	15.79%	21.43%	21.43%	21.43%	0.47%	0.05%	15%	21.43%	21.43%	21.43%	0.39%	0.05%	15%	21.43%	21.43%	21.43%	0.39%	0.05%
30%	20%	26.67%	26.67%	26.67%	0.62%	0.03%	19.05%	26.67%	26.67%	26.67%	0.52%	0.03%	19.05%	26.67%	26.67%	26.67%	0.52%	0.03%
40%	27.27%	35.29%	35.29%	35.29%	0.93%	0.05%	26.09%	35.29%	35.29%	35.29%	0.78%	0.05%	26.09%	35.29%	35.29%	35.29%	0.78%	0.05%
50%	30.44%	38.89%	38.89%	38.89%	1.08%	0.06%	29.17%	38.89%	38.89%	38.89%	0.91%	0.06%	29.17%	38.89%	38.89%	38.89%	0.91%	0.06%
60%	33.33%	42.11%	42.11%	42.11%	1.23%	0.07%	32%	42.11%	42.11%	42.11%	1.04%	0.07%	32%	42.11%	42.11%	42.11%	1.04%	0.07%
70%	0.13%	0.15%	0.15%	0.15%	0.11%	0.08%	0.13%	0.15%	0.14%	0.14%	0.11%	0.08%	0.13%	0.15%	0.14%	0.14%	0.11%	0.08%
80%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%
90%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	45.26%	49.37%	49.37%	49.37%	33.70%	33.36%	44.68%	32.70%	49.37%	49.37%	33.64%	33.36%	44.68%	32.70%	49.37%	49.37%	33.64%	33.36%
AVG100	22.72%	26.47%	26.47%	26.47%	10.45%	10.05%	22.16%	21.47%	26.47%	26.47%	10.38%	10.05%	22.16%	21.47%	26.47%	26.47%	10.38%	10.05%

T416 Query 7 entropy

T416 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.05%	0.05%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30%	0.06%	0.06%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
40%	0.01%	0.01%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
50%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
60%	0.02%	0.02%	11.59%	11.59%	19.05%	5.63%	13.12%	12.31%	11.27%	11.27%	19.05%	5.03%	13.12%	12.31%	11.27%	11.27%	19.05%	5.03%
70%	0.02%	0.02%	0.02%	0.02%	0.04%	0.03%	0.04%	0.02%	0.04%	0.04%	0.07%	0.35%	0.04%	0.02%	0.04%	0.04%	0.07%	0.35%
80%	0.02%	0.02%	0.02%	0.02%	0.02%	0.04%	0.03%	0.02%	0.03%	0.03%	0.04%	0.04%	0.03%	0.02%	0.03%	0.03%	0.04%	0.04%
90%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.02%	0.03%	0.02%	0.02%	0.02%	0.02%	0.02%	0.03%
100%	0%	0%	0%	0%	0%	0%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%
Avg30	0.04%	0.04%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AVG100	0.02%	0.02%	51.16%	51.16%	51.91%	50.57%	51.32%	51.24%	51.14%	51.14%	51.92%	50.55%	51.32%	51.24%	51.14%	51.14%	51.92%	50.55%

T416 Query 8 entropy

T416 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	21.43%	23.08%	37.50%	42.86%	9.09%	1.30%	42.86%	37.50%	37.50%	42.86%	8.82%	1.30%	37.50%	37.50%	33.33%	42.86%	8.82%	1.30%
30%	26.67%	28.57%	44.44%	50%	11.77%	1.73%	50%	44.44%	44.44%	50%	11.43%	1.72%	44.44%	44.44%	40%	50%	11.43%	1.72%
40%	35.29%	37.50%	54.55%	60%	16.67%	2.58%	60%	54.55%	54.55%	60%	16.22%	2.56%	54.55%	54.55%	50%	60%	16.22%	2.56%
50%	38.89%	41.18%	58.33%	63.64%	18.92%	2.99%	63.64%	58.33%	58.33%	63.64%	18.42%	2.98%	58.33%	58.33%	53.85%	63.64%	18.42%	2.98%
60%	42.11%	44.44%	61.54%	66.67%	21.05%	3.40%	66.67%	61.54%	61.54%	66.67%	20.51%	3.39%	61.54%	61.54%	57.14%	66.67%	20.51%	3.39%
70%	0.06%	0.07%	0.08%	0.07%	0.05%	0.10%	0.09%	0.08%	0.08%	0.07%	0.05%	0.10%	0.08%	0.06%	0.07%	0.06%	0.05%	0.10%
80%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
90%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	49.37%	50.55%	60.65%	64.29%	40.29%	34.35%	64.29%	60.65%	60.65%	64.29%	40.08%	34.34%	60.65%	60.65%	57.78%	64.29%	40.08%	34.34%
AVG100	26.45%	27.49%	35.65%	38.33%	17.76%	11.21%	38.33%	35.65%	35.65%	38.33%	17.55%	11.21%	35.65%	35.65%	33.44%	38.33%	17.55%	11.21%

T416 Query 9 entropy

T416 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.23%	0.23%	100%	100%	14.29%	50%	100%	100%	100%	100%	14.29%	50%	100%	100%	100%	100%	14.29%	50%
20%	0.29%	0.29%	100%	100%	33.33%	75%	100%	100%	100%	100%	33.33%	75%	100%	100%	100%	100%	33.33%	75%
30%	0.23%	0.23%	100%	100%	40%	80%	100%	100%	100%	100%	40%	80%	100%	100%	100%	100%	40%	80%
40%	0.33%	0.33%	100%	100%	50%	85.71%	100%	100%	100%	100%	50%	85.71%	100%	100%	100%	100%	50%	85.71%
50%	0.36%	0.36%	100%	100%	53.85%	87.50%	100%	100%	100%	100%	53.85%	87.50%	100%	100%	100%	100%	53.85%	87.50%
60%	0.39%	0.39%	1.68%	1.68%	12.50%	42.11%	100%	5.52%	1.51%	1.51%	12.50%	40%	100%	5.52%	1.51%	1.51%	12.50%	40%
70%	0%	0%	0%	0%	0.43%	2.14%	0.64%	0.29%	0.46%	0.46%	0.33%	1.20%	0.64%	0.29%	0.46%	0.46%	0.33%	1.20%
80%	0%	0%	0%	0%	0.06%	0.06%	0.23%	0.23%	0.23%	0.23%	0.06%	0.42%	0.23%	0.23%	0.23%	0.23%	0.06%	0.42%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.04%	0.07%	0%	0%	0%	0%	0.04%	0.07%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%
Avg30	0.25%	0.25%	100.00%	100.00%	29.21%	68.33%	100.00%	100.00%	100.00%	100.00%	29.21%	68.33%	100.00%	100.00%	100.00%	100.00%	29.21%	68.33%
AVG100	0.18%	0.18%	50.17%	50.17%	20.45%	42.25%	60.09%	50.60%	50.22%	50.22%	20.44%	41.99%	60.09%	50.60%	50.22%	50.22%	20.44%	41.99%

T416 Query 10 entropy

T416 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	21.43%	30%	30%	42.86%	0.17%	0.10%	21.43%	30%	30%	42.86%	0.17%	0.10%	20%	30%	30%	42.86%	0.17%	0.10%
30%	26.67%	36.36%	36.36%	50%	0.23%	0.14%	26.67%	36.36%	36.36%	50%	0.23%	0.14%	25%	36.36%	36.36%	50%	0.23%	0.14%
40%	35.29%	46.15%	46.15%	60%	0.34%	0.21%	35.29%	46.15%	46.15%	60%	0.34%	0.21%	33.33%	46.15%	46.15%	60%	0.34%	0.21%
50%	38.89%	50%	50%	63.64%	0.39%	0.24%	38.89%	50%	50%	63.64%	0.39%	0.24%	36.84%	50%	50%	63.64%	0.39%	0.24%
60%	42.11%	53.33%	53.33%	66.67%	0.45%	0.28%	42.11%	53.33%	53.33%	66.67%	0.45%	0.28%	40%	53.33%	53.33%	66.67%	0.45%	0.28%
70%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.05%	0.05%
80%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.04%	0.04%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	49.37%	55.45%	55.45%	64.29%	33.46%	33.41%	49.37%	55.45%	55.45%	64.29%	33.46%	33.41%	48.33%	55.45%	55.45%	64.29%	33.46%	33.41%
AVG100	26.44%	31.59%	31.59%	38.32%	10.16%	10.10%	26.44%	31.59%	31.59%	38.32%	10.16%	10.10%	25.52%	31.59%	31.59%	38.32%	10.17%	10.11%

T416 Query 1 idf

T416 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	37.50%	60%	50%	75%	30%	60%	37.50%	50%	50%	75%	30%	60%	50%	75%	75%	75%	30%	60%
30%	44.44%	66.67%	57.14%	80%	36.36%	66.67%	44.44%	57.14%	57.14%	80%	36.36%	66.67%	57.14%	80%	80%	80%	36.36%	66.67%
40%	54.55%	75%	66.67%	85.71%	46.15%	75%	54.55%	66.67%	66.67%	85.71%	46.15%	75%	66.67%	85.71%	85.71%	85.71%	46.15%	75%
50%	58.33%	77.78%	70%	87.50%	50%	77.78%	58.33%	70%	70%	87.50%	50%	77.78%	70%	87.50%	87.50%	87.50%	50%	77.78%
60%	61.54%	80%	72.73%	88.89%	53.33%	80%	61.54%	72.73%	72.73%	88.89%	53.33%	80%	72.73%	88.89%	88.89%	88.89%	53.33%	80%
70%	33.33%	58.82%	20%	71.43%	1.09%	55.56%	34.48%	20%	20%	71.43%	1.09%	55.56%	33.33%	71.43%	71.43%	71.43%	0.80%	55.56%
80%	29.73%	44%	1.99%	45.83%	1.19%	57.90%	29.73%	1.99%	1.99%	45.83%	1.19%	57.90%	32.35%	42.31%	42.31%	45.83%	0.87%	57.90%
90%	16.46%	18.57%	0.28%	20.97%	0.22%	0.23%	16.25%	0.28%	0.28%	20.97%	0.22%	0.23%	18.31%	18.31%	18.31%	20.97%	0.20%	0.23%
100%	0.22%	0.25%	0.24%	0.23%	0.19%	0.12%	0.21%	0.24%	0.24%	0.23%	0.19%	0.12%	0.19%	0.24%	0.24%	0.21%	0.18%	0.11%
Avg30	60.65%	75.56%	69.05%	85.00%	55.45%	75.56%	60.65%	69.05%	69.05%	85.00%	55.45%	75.56%	69.05%	85.00%	85.00%	85.00%	55.45%	75.56%
AVG100	43.61%	58.11%	43.90%	65.56%	31.85%	57.32%	43.70%	43.90%	43.90%	65.56%	31.85%	57.32%	50.07%	64.94%	64.94%	65.55%	31.79%	57.32%

T416 Query 2 idf

T416 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	33.33%	75%	75%	75%	100%	75%	75%	75%	75%	75%	100%	75%	75%	75%	75%	75%	100%	75%
30%	30.77%	80%	80%	80%	100%	80%	80%	80%	80%	80%	100%	80%	80%	80%	80%	80%	100%	80%
40%	40%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%	85.71%	85.71%	85.71%	85.71%	100%	85.71%
50%	43.75%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%	87.50%	87.50%	87.50%	87.50%	100%	87.50%
60%	47.06%	88.89%	88.89%	88.89%	8.08%	88.89%	88.89%	88.89%	88.89%	88.89%	8%	88.89%	88.89%	88.89%	88.89%	88.89%	8%	88.89%
70%	52.63%	66.67%	66.67%	66.67%	3.95%	66.67%	76.92%	66.67%	66.67%	66.67%	3.88%	66.67%	76.92%	66.67%	66.67%	66.67%	3.88%	66.67%
80%	55%	20.76%	18.33%	18.33%	1.55%	20.76%	78.57%	15.49%	16.92%	16.92%	1.53%	20.76%	78.57%	15.49%	16.92%	16.92%	1.53%	20.76%
90%	24.53%	0.49%	0.49%	0.49%	1.43%	9.22%	50%	8.39%	9.09%	9.09%	1.42%	8.73%	50%	8.39%	9.09%	9.09%	1.42%	8.73%
100%	0%	0%	0%	0%	0.52%	0.57%	43.75%	0.52%	0.49%	0.49%	0.50%	0.56%	43.75%	0.52%	0.49%	0.49%	0.50%	0.56%
Avg30	54.70%	85.00%	85.00%	85.00%	100.00%	85.00%	85.00%	85.00%	85.00%	85.00%	100.00%	85.00%	85.00%	85.00%	85.00%	85.00%	100.00%	85.00%
AVG100	42.71%	60.50%	60.26%	60.26%	51.55%	61.43%	76.63%	60.82%	61.03%	61.03%	51.53%	61.38%	76.63%	60.82%	61.03%	61.03%	51.53%	61.38%

T416 Query 3 idf

T416 Q3	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.98%	0.98%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	1.45%	1.45%	100%	100%	75%	75%	100%	100%	100%	100%	75%	75%	100%	100%	100%	100%	75%	75%
30%	1.79%	1.79%	100%	100%	80%	80%	100%	100%	100%	100%	80%	80%	100%	100%	100%	100%	80%	80%
40%	0.67%	0.67%	100%	100%	85.71%	85.71%	100%	100%	100%	100%	85.71%	85.71%	100%	100%	100%	100%	85.71%	85.71%
50%	0.72%	0.72%	100%	100%	87.50%	87.50%	100%	100%	100%	100%	87.50%	87.50%	100%	100%	100%	100%	87.50%	87.50%
60%	0.79%	0.79%	88.89%	88.89%	88.89%	88.89%	66.67%	61.54%	88.89%	88.89%	88.89%	88.89%	66.67%	61.54%	88.89%	88.89%	88.89%	88.89%
70%	0.90%	0.90%	62.50%	62.50%	58.82%	55.56%	66.67%	66.67%	62.50%	62.50%	58.82%	55.56%	66.67%	66.67%	62.50%	62.50%	58.82%	55.56%
80%	0.93%	0.93%	55%	55%	61.11%	57.90%	45.83%	57.90%	55%	55%	61.11%	57.90%	45.83%	57.90%	55%	55%	61.11%	57.90%
90%	0.89%	0.89%	41.94%	41.94%	22.41%	44.83%	50%	40.63%	41.94%	41.94%	22.41%	44.83%	50%	40.63%	41.94%	41.94%	22.41%	44.83%
100%	0.87%	0.87%	1.02%	1.02%	1.70%	20.90%	1.24%	1.59%	0.69%	0.69%	1.70%	20.90%	1.24%	1.59%	0.69%	0.69%	1.70%	20.90%
Avg30	1.41%	1.41%	100.00%	100.00%	85.00%	85.00%	100.00%	100.00%	100.00%	100.00%	85.00%	85.00%	100.00%	100.00%	100.00%	100.00%	85.00%	85.00%
AVG100	1.00%	1.00%	74.93%	74.93%	66.12%	69.63%	73.04%	72.83%	74.90%	74.90%	66.12%	69.63%	73.04%	72.83%	74.90%	74.90%	66.12%	69.63%

T416 Query 4 idf

T416 Q4	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	60%	100%	100%	100%	10.71%	100%	60%	100%	100%	100%	10.71%	100%	60%	100%	100%	100%	10.71%	100%
30%	66.67%	100%	100%	100%	13.79%	100%	66.67%	100%	100%	100%	13.79%	100%	66.67%	100%	100%	100%	13.79%	100%
40%	75%	100%	100%	100%	19.36%	100%	75%	100%	100%	100%	19.36%	100%	75%	100%	100%	100%	19.36%	100%
50%	77.78%	100%	100%	100%	21.88%	100%	77.78%	100%	100%	100%	21.88%	100%	77.78%	100%	100%	100%	21.88%	100%
60%	80%	100%	100%	100%	24.24%	100%	80%	100%	100%	100%	24.24%	100%	80%	100%	100%	100%	24.24%	100%
70%	12.99%	27.78%	27.78%	27.78%	0.93%	0.33%	12.99%	27.78%	27.78%	27.78%	0.93%	0.33%	12.99%	27.78%	27.78%	27.78%	0.93%	0.33%
80%	0.40%	0.40%	0.40%	0.40%	0.38%	0.33%	0.40%	0.40%	0.40%	0.40%	0.38%	0.33%	0.40%	0.40%	0.40%	0.40%	0.38%	0.33%
90%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%	0.41%	0.40%	0.40%	0.40%	0.05%	0.05%
100%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%
Avg30	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%	75.56%	100.00%	100.00%	100.00%	41.50%	100.00%
AVG100	47.32%	62.86%	62.86%	62.86%	19.14%	60.08%	47.32%	62.86%	62.86%	62.86%	19.14%	60.08%	47.32%	62.86%	62.86%	62.86%	19.14%	60.08%

T416 Query 5 idf

T416 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.16%	0.16%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.36%	0.36%	75%	75%	42.86%	100%	60%	100%	60%	60%	42.86%	100%	60%	100%	60%	100%	42.86%	100%
30%	0.46%	0.46%	66.67%	66.67%	50%	100%	66.67%	100%	66.67%	66.67%	50%	100%	66.67%	100%	66.67%	100%	50%	100%
40%	0.29%	0.29%	75%	75%	60%	0.20%	42.86%	100%	75%	75%	60%	0.20%	42.86%	100%	75%	75%	60%	0.20%
50%	0.34%	0.34%	77.78%	77.78%	63.64%	0.24%	46.67%	100%	77.78%	77.78%	63.64%	0.24%	46.67%	100%	77.78%	77.78%	63.64%	0.24%
60%	0.38%	0.38%	80%	80%	66.67%	0.27%	50%	100%	80%	80%	66.67%	0.27%	50%	100%	80%	80%	66.67%	0.27%
70%	0.42%	0.42%	83.33%	83.33%	17.54%	0.34%	55.56%	50%	71.43%	71.43%	17.54%	0.34%	55.56%	50%	71.43%	71.43%	17.54%	0.34%
80%	0.43%	0.43%	73.33%	73.33%	18.97%	0.37%	57.90%	30.56%	26.83%	26.83%	18.97%	0.37%	57.90%	30.56%	26.83%	26.83%	18.97%	0.37%
90%	0.45%	0.45%	27.66%	27.66%	0.77%	0.34%	3.93%	0.46%	1.47%	1.47%	0.75%	0.27%	3.93%	0.46%	1.47%	1.47%	0.75%	0.27%
100%	0%	0%	1.85%	1.85%	0.04%	0.36%	1.32%	0.40%	0.34%	0.34%	0.04%	0.29%	1.32%	0.40%	0.34%	0.34%	0.04%	0.29%
Avg30	0.33%	0.33%	80.56%	80.56%	64.29%	100.00%	75.56%	100.00%	75.56%	75.56%	64.29%	100.00%	75.56%	100.00%	75.56%	100.00%	64.29%	100.00%
AVG100	0.33%	0.33%	66.06%	66.06%	42.05%	30.21%	48.49%	68.14%	55.95%	55.95%	42.05%	30.20%	48.49%	68.14%	55.95%	63.28%	42.05%	30.20%

T416 Query 6 idf

T416 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	100%	100%	100%	100%	100%	100%	100%	50%	100%	100%	100%	100%	100%	50%	100%	100%	100%	100%
20%	15.79%	21.43%	21.43%	21.43%	0.44%	0.05%	13.64%	21.43%	21.43%	21.43%	0.38%	0.05%	13.64%	21.43%	21.43%	21.43%	0.38%	0.05%
30%	20%	26.67%	26.67%	26.67%	0.59%	0.03%	17.39%	26.67%	26.67%	26.67%	0.50%	0.03%	17.39%	26.67%	26.67%	26.67%	0.50%	0.03%
40%	27.27%	35.29%	35.29%	35.29%	0.88%	0.05%	24%	35.29%	35.29%	35.29%	0.75%	0.05%	24%	35.29%	35.29%	35.29%	0.75%	0.05%
50%	30.44%	38.89%	38.89%	38.89%	1.02%	0.06%	26.92%	38.89%	38.89%	38.89%	0.87%	0.06%	26.92%	38.89%	38.89%	38.89%	0.87%	0.06%
60%	33.33%	42.11%	42.11%	42.11%	1.17%	0.07%	29.63%	42.11%	42.11%	42.11%	0.99%	0.07%	29.63%	42.11%	42.11%	42.11%	0.99%	0.07%
70%	0.13%	0.15%	0.14%	0.14%	0.11%	0.08%	0.12%	0.15%	0.14%	0.14%	0.11%	0.08%	0.12%	0.15%	0.14%	0.14%	0.11%	0.08%
80%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%	0.09%	0.09%	0.09%	0.09%	0.03%	0.09%
90%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%	0.10%	0.10%	0.10%	0.10%	0.03%	0.10%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	45.26%	49.37%	49.37%	49.37%	33.68%	33.36%	43.68%	32.70%	49.37%	49.37%	33.62%	33.36%	43.68%	32.70%	49.37%	49.37%	33.62%	33.36%
AVG100	22.72%	26.47%	26.47%	26.47%	10.43%	10.05%	21.19%	21.47%	26.47%	26.47%	10.37%	10.05%	21.19%	21.47%	26.47%	26.47%	10.37%	10.05%

T416 Query 7 idf

T416 Q7	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	0.05%	0.05%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
30%	0.06%	0.06%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
40%	0.01%	0.01%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
50%	0.02%	0.02%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
60%	0.02%	0.02%	12.31%	12.31%	19.51%	5.93%	14.04%	12.90%	11.94%	11.94%	19.51%	5.48%	14.04%	12.90%	11.94%	11.94%	19.51%	5.48%
70%	0.02%	0.02%	0.02%	0.02%	0.04%	0.03%	0.04%	0.02%	0.04%	0.04%	0.07%	0.34%	0.04%	0.02%	0.04%	0.04%	0.07%	0.34%
80%	0.02%	0.02%	0.02%	0.02%	0.02%	0.04%	0.03%	0.02%	0.03%	0.03%	0.04%	0.04%	0.03%	0.02%	0.03%	0.03%	0.04%	0.04%
90%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.02%	0.03%	0.02%	0.02%	0.02%	0.02%	0.02%	0.03%
100%	0%	0%	0%	0%	0%	0%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%	0.02%	0.02%	0.02%	0.02%	0.01%	0.01%
Avg30	0.04%	0.04%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
AVG100	0.02%	0.02%	51.24%	51.24%	51.96%	50.60%	51.41%	51.30%	51.20%	51.20%	51.97%	50.59%	51.41%	51.30%	51.20%	51.20%	51.97%	50.59%

T416 Query 8 idf

T416 Q8	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	21.43%	23.08%	37.50%	33.33%	9.09%	1.30%	42.86%	37.50%	37.50%	33.33%	8.82%	1.29%	37.50%	33.33%	33.33%	33.33%	8.82%	1.29%
30%	26.67%	28.57%	44.44%	40%	11.77%	1.72%	50%	44.44%	44.44%	40%	11.43%	1.72%	44.44%	40%	40%	40%	11.43%	1.72%
40%	35.29%	37.50%	54.55%	50%	16.67%	2.56%	60%	54.55%	54.55%	50%	16.22%	2.55%	54.55%	50%	50%	50%	16.22%	2.55%
50%	38.89%	41.18%	58.33%	53.85%	18.92%	2.98%	63.64%	58.33%	58.33%	53.85%	18.42%	2.97%	58.33%	53.85%	53.85%	53.85%	18.42%	2.97%
60%	42.11%	44.44%	61.54%	57.14%	21.05%	3.39%	66.67%	61.54%	61.54%	57.14%	20.51%	3.38%	61.54%	57.14%	57.14%	57.14%	20.51%	3.38%
70%	0.06%	0.06%	0.08%	0.07%	0.05%	0.11%	0.09%	0.08%	0.08%	0.07%	0.05%	0.11%	0.08%	0.07%	0.07%	0.06%	0.05%	0.10%
80%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
90%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	49.37%	50.55%	60.65%	57.78%	40.29%	34.34%	64.29%	60.65%	60.65%	57.78%	40.08%	34.34%	60.65%	57.78%	57.78%	57.78%	40.08%	34.34%
AVG100	26.45%	27.49%	35.65%	33.44%	17.76%	11.21%	38.33%	35.65%	35.65%	33.44%	17.55%	11.20%	35.65%	33.44%	33.44%	33.44%	17.55%	11.20%

T416 Query 9 idf

T416 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.23%	0.23%	100%	100%	12.50%	50%	100%	100%	100%	100%	12.50%	50%	100%	100%	100%	100%	12.50%	50%
20%	0.29%	0.29%	100%	100%	30%	75%	100%	100%	100%	100%	30%	75%	100%	100%	100%	100%	30%	75%
30%	0.23%	0.23%	100%	100%	36.36%	80%	100%	100%	100%	100%	36.36%	80%	100%	100%	100%	100%	36.36%	80%
40%	0.33%	0.33%	100%	100%	46.15%	85.71%	100%	100%	100%	100%	46.15%	85.71%	100%	100%	100%	100%	46.15%	85.71%
50%	0.36%	0.36%	100%	100%	50%	87.50%	100%	100%	100%	100%	50%	87.50%	100%	100%	100%	100%	50%	87.50%
60%	0.39%	0.39%	5.03%	5.03%	3.64%	40%	100%	5.56%	4.73%	4.73%	2.82%	38.10%	100%	5.56%	4.73%	4.73%	2.82%	38.10%
70%	0%	0%	0%	0%	0.48%	2.17%	0.64%	0.29%	0.46%	0.46%	0.37%	1.22%	0.64%	0.29%	0.46%	0.46%	0.37%	1.22%
80%	0%	0%	0%	0%	0.06%	0.06%	0.23%	0.23%	0.23%	0.23%	0.06%	0.42%	0.23%	0.23%	0.23%	0.23%	0.06%	0.42%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.07%	0%	0%	0%	0%	0.03%	0.07%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%
Avg30	0.25%	0.25%	100.00%	100.00%	26.29%	68.33%	100.00%	100.00%	100.00%	100.00%	26.29%	68.33%	100.00%	100.00%	100.00%	100.00%	26.29%	68.33%
AVG100	0.18%	0.18%	50.50%	50.50%	17.92%	42.04%	60.09%	50.61%	50.54%	50.54%	17.83%	41.81%	60.09%	50.61%	50.54%	50.54%	17.83%	41.81%

T416 Query 10 idf

T416 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	21.43%	30%	30%	37.50%	0.17%	0.10%	21.43%	30%	30%	37.50%	0.17%	0.10%	20%	25%	25%	37.50%	0.17%	0.10%
30%	26.67%	36.36%	36.36%	44.44%	0.23%	0.14%	26.67%	36.36%	36.36%	44.44%	0.23%	0.14%	25%	30.77%	30.77%	44.44%	0.23%	0.14%
40%	35.29%	46.15%	46.15%	54.55%	0.34%	0.21%	35.29%	46.15%	46.15%	54.55%	0.34%	0.21%	33.33%	40%	40%	54.55%	0.34%	0.21%
50%	38.89%	50%	50%	58.33%	0.39%	0.24%	38.89%	50%	50%	58.33%	0.39%	0.24%	36.84%	43.75%	43.75%	58.33%	0.39%	0.24%
60%	42.11%	53.33%	53.33%	61.54%	0.45%	0.28%	42.11%	53.33%	53.33%	61.54%	0.45%	0.28%	40%	47.06%	47.06%	61.54%	0.45%	0.28%
70%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.05%	0.05%
80%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.03%	0.03%	0%	0%	0%	0%	0.04%	0.04%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	49.37%	55.45%	55.45%	60.65%	33.46%	33.41%	49.37%	55.45%	55.45%	60.65%	33.46%	33.41%	48.33%	51.92%	51.92%	60.65%	33.46%	33.41%
AVG100	26.44%	31.59%	31.59%	35.64%	10.16%	10.10%	26.44%	31.59%	31.59%	35.64%	10.16%	10.10%	25.52%	28.66%	28.66%	35.64%	10.17%	10.11%

T431 Query 1

T431 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	14.42%	15.31%	19.74%	19.74%	16.85%	16.30%	18.99%	20.00%	19.74%	19.74%	16.85%	16.30%	16.85%	20%	19.74%	19.74%	16.85%	16.30%
20%	15.96%	15.39%	16.76%	16.76%	15.96%	16.95%	17%	16.39%	16.76%	16.76%	15.96%	16.58%	15.96%	16.39%	16.76%	16.76%	15.96%	16.95%
30%	18.03%	15.88%	16.42%	16.42%	16.06%	14.87%	16.99%	16.06%	16.42%	16.42%	16.06%	14.87%	16.99%	16.06%	16.42%	16.42%	16.06%	14.87%
40%	11.01%	10.69%	12.47%	12.47%	12.09%	14.15%	12.04%	12.34%	12.47%	12.47%	12.09%	14.15%	12.04%	12.34%	12.47%	12.47%	12.09%	14.15%
50%	10.69%	10.91%	9.97%	9.97%	9.07%	12.21%	10.51%	9.56%	9.97%	9.97%	9.07%	12.13%	10.54%	9.56%	9.97%	9.97%	9.06%	12.21%
60%	11.32%	10.53%	8.47%	8.47%	8.52%	10.61%	8.43%	8.27%	8.47%	8.47%	8.52%	10.56%	8.44%	8.27%	8.47%	8.47%	8.52%	10.61%
70%	8.84%	7.23%	6.78%	6.78%	7.06%	7.55%	6.79%	6.66%	6.78%	6.78%	7.06%	7.53%	6.82%	6.66%	6.78%	6.78%	7.06%	7.55%
80%	7.12%	6.30%	6.55%	6.55%	6.47%	7.02%	6.32%	6.24%	6.55%	6.55%	6.47%	7.02%	6.33%	6.24%	6.55%	6.55%	6.47%	7.02%
90%	5.65%	5.79%	3.41%	3.41%	3.55%	4.87%	3.06%	2.95%	3.41%	3.41%	3.54%	4.87%	3.05%	2.95%	3.41%	3.41%	3.55%	4.87%
100%	0.34%	0.33%	0.34%	0.34%	0.36%	0.36%	0.29%	0.42%	0.26%	0.26%	0.31%	0.28%	0.29%	0.42%	0.26%	0.26%	0.31%	0.28%
Avg30	16.14%	15.53%	17.64%	17.64%	16.29%	16.04%	17.61%	17.48%	17.64%	17.64%	16.29%	15.91%	16.60%	17.48%	17.64%	17.64%	16.29%	16.04%
AVG100	10.34%	9.84%	10.09%	10.09%	9.60%	10.49%	10.03%	9.89%	10.08%	10.08%	9.59%	10.43%	9.73%	9.89%	10.08%	10.08%	9.59%	10.48%

T431 Query 2

T431 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	19.48%	19.74%	15.96%	22.73%	31.92%	17.24%	18.07%	18.99%	15.96%	22.73%	31.92%	17.24%	22.39%	25.42%	22.39%	22.73%	33.33%	17.44%
20%	23.08%	22.56%	20.83%	18.07%	19.36%	17.34%	21.13%	17.75%	20.83%	17.86%	19.36%	16.95%	17.34%	17.34%	17.54%	17.34%	19.48%	16.39%
30%	18.64%	18.57%	24.04%	15.02%	19.47%	14.87%	23.66%	20.47%	24.04%	14.92%	19.47%	14.52%	15.02%	16.79%	14.67%	15.23%	19.64%	13.88%
40%	16.91%	16.21%	20.14%	12.32%	17.51%	12.32%	20.07%	19.87%	20.14%	12.17%	17.51%	11.99%	13.56%	15.37%	12.17%	12.97%	17.93%	11.05%
50%	16.74%	16.34%	19.79%	10.50%	15.39%	8.79%	19.79%	18.59%	19.79%	10.44%	15.39%	8.92%	12.21%	13.38%	10.22%	12.07%	15.61%	8.08%
60%	16.04%	15.87%	17.84%	9.51%	11.57%	6.14%	17.80%	15.11%	17.84%	9.38%	11.57%	6.14%	8.98%	9.83%	9.34%	9.35%	11.76%	5.68%
70%	12.21%	12.02%	15.16%	8.68%	11.20%	6.40%	15.03%	12.15%	15.16%	8.73%	11.20%	6.40%	8.31%	9.47%	8.73%	8.82%	11.23%	5.92%
80%	9.58%	9.95%	12.27%	7.98%	8.53%	6.16%	12.07%	9.80%	12.22%	8.02%	8.53%	6.11%	7.49%	8.25%	8.04%	7.88%	8.55%	5.74%
90%	5.40%	5.28%	8.38%	5.51%	5.88%	4.84%	9.14%	6.39%	8.33%	5.51%	5.88%	4.54%	6.20%	6.90%	5.51%	5.46%	5.71%	4.34%
100%	4.37%	4.36%	3.22%	0.47%	0.51%	0.45%	3.74%	4.38%	2.51%	0.25%	0.29%	0.25%	0.25%	4.34%	0.25%	0.25%	0.29%	0.25%
Avg30	20.40%	20.29%	20.28%	18.61%	23.58%	16.48%	20.95%	19.07%	20.28%	18.50%	23.58%	16.24%	18.25%	19.85%	18.20%	18.43%	24.15%	15.91%
AVG100	14.24%	14.09%	15.76%	11.08%	14.13%	9.45%	16.05%	14.35%	15.68%	11.00%	14.11%	9.31%	11.18%	12.71%	10.89%	11.21%	14.35%	8.88%

T431 Query 3

T431 Q3	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	15.15%	18%	17.05%	17.05%	22.39%	15.31%	17.44%	19.23%	17.05%	17.05%	22.39%	15.31%	15.31%	21.43%	25.86%	25.86%	21.13%	20%
20%	20.69%	20.13%	22.06%	22.06%	23.81%	21.13%	22.73%	22.06%	22.06%	22.06%	23.81%	21.43%	21.43%	18.99%	15.96%	15.96%	19.48%	17.14%
30%	23.91%	23.78%	26.04%	26.04%	24.31%	19.47%	25.58%	23%	26.04%	26.04%	24.31%	19.38%	19.38%	20.28%	18.33%	18.33%	20.09%	15.23%
40%	24.38%	19.93%	24.69%	24.69%	22.35%	13.02%	24.48%	23.98%	24.69%	24.69%	22.35%	13.11%	17.35%	20.21%	17.51%	17.51%	21.38%	12.83%
50%	18.50%	18.18%	19.27%	19.27%	19.73%	7.20%	19.73%	19.89%	19.22%	19.22%	19.73%	7.16%	17.92%	19.07%	17.87%	17.87%	19.73%	12.23%
60%	17.87%	17.55%	18.05%	18.05%	17.87%	5.23%	18.28%	17.69%	18.02%	18.02%	17.87%	5.29%	16.18%	17.66%	16.86%	16.86%	17.80%	9.24%
70%	15.88%	14.75%	16.48%	16.48%	14.31%	1.50%	15.88%	14.39%	16.43%	16.43%	14.31%	1.50%	13.45%	15.81%	14.37%	14.37%	14.49%	8.85%
80%	14.13%	11.32%	13.20%	13.20%	10.23%	1.58%	12.97%	13.07%	13.14%	13.14%	10.23%	1.58%	10.47%	13.05%	10.71%	10.71%	10.72%	6.90%
90%	9.26%	6.73%	9.39%	9.39%	5.39%	1.68%	9.99%	8.20%	9.28%	9.28%	5.39%	1.68%	7.62%	9.19%	7.26%	7.26%	5.41%	2.65%
100%	2.64%	1.13%	3.19%	3.19%	1.03%	1.16%	3.44%	1.15%	3.13%	3.13%	1.06%	1.15%	2.20%	1.15%	1.20%	1.20%	1.05%	0.23%
Avg30	19.92%	20.59%	21.71%	21.71%	23.50%	18.63%	21.92%	21.56%	21.71%	21.71%	23.50%	18.71%	18.71%	20.23%	20.05%	20.05%	20.23%	17.46%
AVG100	16.24%	15.14%	16.94%	16.94%	16.14%	8.73%	17.05%	16.31%	16.90%	16.90%	16.14%	8.76%	14.13%	15.68%	14.59%	14.59%	15.13%	10.53%

T431 Query 4

T431 Q4	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	22.39%	14.02%	13.89%	30.00%	17.65%	16.67%	15.31%	13.89%	13.89%	30%	17.65%	16.67%	17.65%	22.39%	22.39%	22.39%	17.44%	29.41%
20%	8.52%	11.95%	12.05%	14%	12.66%	20.27%	14.42%	11.95%	11.95%	14.09%	12.66%	20.27%	12.66%	15%	15%	15%	18.52%	17.14%
30%	6.41%	13.46%	13.58%	8.93%	6.64%	15.28%	10.26%	13.42%	13.42%	8.93%	6.62%	15.28%	6.62%	12.83%	12.83%	12.83%	15.94%	15.33%
40%	7.48%	9.80%	9.82%	9.06%	6.41%	2.40%	9.25%	9.58%	9.58%	9.01%	6.40%	2.38%	8.77%	10.87%	10.87%	10.87%	16.08%	2.38%
50%	7.61%	10.31%	10.31%	8.76%	6.37%	2.72%	8.59%	10.08%	10.08%	8.70%	6.36%	2.71%	8.57%	9.16%	9.16%	9.16%	14.57%	2.71%
60%	7.95%	9.59%	9.59%	9.16%	6.15%	2.51%	8.41%	9.42%	9.42%	9.10%	6.14%	2.50%	9.05%	9.35%	9.35%	9.35%	12.55%	2.50%
70%	8.11%	9.06%	9.06%	8.04%	5.76%	2.88%	8.21%	8.93%	8.93%	8.01%	5.76%	2.86%	9.24%	9.83%	9.83%	9.83%	12.67%	2.87%
80%	7.88%	8.45%	8.45%	6.41%	4.69%	3.19%	7.77%	8.35%	8.35%	6.39%	4.69%	3.17%	8.00%	8.12%	8.12%	8.12%	10.86%	3.18%
90%	6.07%	5.07%	5.07%	4.46%	4.11%	3.49%	5.70%	5.04%	5.04%	4.45%	4.11%	3.46%	4.98%	4.77%	4.77%	4.77%	6.12%	3.47%
100%	2.91%	2.92%	2.92%	2.92%	0.41%	2.90%	2.89%	2.91%	2.91%	2.91%	0.41%	2.89%	2.91%	2.91%	2.91%	2.91%	2.90%	2.89%
Avg30	12.44%	13.14%	13.17%	17.67%	12.31%	17.41%	13.33%	13.09%	13.09%	17.67%	12.31%	17.41%	12.31%	16.74%	16.74%	16.74%	17.30%	20.63%
AVG100	8.53%	9.46%	9.47%	10.18%	7.09%	7.23%	9.08%	9.36%	9.36%	10.16%	7.08%	7.22%	8.84%	10.52%	10.52%	10.52%	12.76%	8.19%

T431 Query 5

T431 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	18.75%	19.48%	17.86%	18.52%	21.13%	24.19%	18.29%	22.73%	17.65%	18.29%	21.13%	24.19%	22.73%	15.79%	17.65%	17.65%	18.07%	20.27%
20%	18.99%	21.13%	18.63%	21.43%	17.75%	22.39%	18.52%	12.88%	18.18%	20.98%	17.75%	22.39%	12.88%	20%	20.55%	20.55%	19.23%	19.61%
30%	17.96%	20.47%	20.18%	19.82%	14.67%	19.56%	17.81%	9.42%	19.73%	19.47%	14.67%	19.47%	9.42%	16%	19.73%	19.73%	15.12%	19.91%
40%	18.27%	18.44%	19.03%	18.73%	14.60%	13.00%	18.32%	9.52%	18.44%	18.15%	14.60%	13.00%	18.67%	14.25%	19.03%	19.03%	15.65%	9.13%
50%	15.26%	17.01%	16.67%	16.12%	13.29%	5.23%	14.26%	9.75%	16.23%	15.71%	13.29%	5.21%	16.37%	13.07%	15.51%	15.51%	13.88%	5.22%
60%	13.71%	14.11%	14.15%	14.08%	12.18%	2.73%	12.16%	10.34%	13.76%	13.69%	12.18%	2.72%	12.47%	11.54%	13.78%	13.78%	12.45%	2.71%
70%	11.94%	12.21%	12.08%	12.62%	12.34%	3.08%	10.80%	10.21%	11.73%	12.24%	12.34%	3.07%	11.34%	10.79%	12.28%	12.28%	12.25%	3.08%
80%	10.68%	10.75%	9.86%	10.21%	10.00%	3.37%	9.36%	8.73%	9.55%	9.87%	10%	3.36%	9.58%	9.28%	9.96%	9.96%	10.07%	3.35%
90%	6.48%	5.72%	5.07%	6.45%	6.08%	3.65%	5.44%	4.60%	4.78%	6.14%	6.08%	3.63%	5.67%	6.08%	6.20%	6.20%	6.08%	3.54%
100%	3.06%	3.09%	2.34%	2.34%	2.23%	0.51%	1.76%	3.07%	1.76%	1.76%	1.77%	0.49%	1.76%	3.07%	1.76%	1.76%	1.77%	1.75%
Avg30	18.57%	20.36%	18.89%	19.92%	17.85%	22.05%	18.21%	15.01%	18.52%	19.58%	17.85%	22.02%	15.01%	17.26%	19.31%	19.31%	17.47%	19.93%
AVG100	13.51%	14.24%	13.59%	14.03%	12.43%	9.77%	12.67%	10.12%	13.18%	13.63%	12.38%	9.75%	12.09%	11.99%	13.65%	13.65%	12.46%	8.86%

T431 Query 6

T431 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0%	0%	1.98%	1.98%	10.00%	7.81%	3.98%	11.36%	3.98%	3.98%	13.89%	11.45%	51.72%	25.86%	48.39%	48.39%	41.67%	31.25%
20%	0%	0%	1.18%	1.18%	0.75%	0.96%	1.93%	0.61%	1.98%	1.98%	3.66%	0.61%	31.25%	30.93%	31.25%	31.25%	34.88%	31.25%
30%	0%	0%	0.29%	0.29%	0.26%	0.26%	1.22%	0.44%	1.17%	1.17%	1.80%	0.44%	0.83%	0.80%	0.80%	0.80%	2.29%	0.80%
40%	0%	0%	0%	0%	0.16%	0.16%	0.66%	0.35%	0.71%	0.71%	1.08%	0.35%	0.41%	0.41%	0.41%	0.41%	0.55%	0.41%
50%	0%	0%	0%	0%	0%	0%	0.32%	0.26%	0.34%	0.34%	0.61%	0.26%	0.28%	0.28%	0.28%	0.28%	0.42%	0.28%
60%	0%	0%	0%	0%	0%	0%	0.25%	0.22%	0.25%	0.25%	0.31%	0.21%	0.22%	0.22%	0.22%	0.22%	0.29%	0.22%
70%	0%	0%	0%	0%	0%	0%	0.18%	0.18%	0.19%	0.19%	0.19%	0.19%	0.18%	0.18%	0.18%	0.18%	0.20%	0.19%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.00%	0.00%	1.15%	1.15%	3.67%	3.01%	2.38%	4.14%	2.38%	2.38%	6.45%	4.17%	27.93%	19.20%	26.81%	26.81%	26.28%	21.10%
AVG100	0.00%	0.00%	0.35%	0.35%	1.12%	0.92%	0.86%	1.34%	0.86%	0.86%	2.15%	1.35%	8.49%	5.87%	8.15%	8.15%	8.03%	6.44%

T431 Query 7

T431 Q7	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	14.42%	15.31%	19.74%	19.74%	16.85%	16.30%	18.99%	20.00%	19.74%	19.74%	16.85%	16.30%	16.85%	20%	19.74%	19.74%	16.85%	16.30%
20%	15.96%	15.39%	16.76%	16.76%	15.96%	16.95%	17%	16.39%	16.76%	16.76%	15.96%	16.58%	15.96%	16.39%	16.76%	16.76%	15.96%	16.95%
30%	18.03%	15.88%	16.42%	16.42%	16.06%	14.87%	16.99%	16.06%	16.42%	16.42%	16.06%	14.87%	16.99%	16.06%	16.42%	16.42%	16.06%	14.87%
40%	11.01%	10.69%	12.47%	12.47%	12.09%	14.15%	12.04%	12.34%	12.47%	12.47%	12.09%	14.15%	12.04%	12.34%	12.47%	12.47%	12.09%	14.15%
50%	10.69%	10.91%	9.97%	9.97%	9.07%	12.21%	10.51%	9.56%	9.97%	9.97%	9.07%	12.13%	10.54%	9.56%	9.97%	9.97%	9.06%	12.21%
60%	11.32%	10.53%	8.47%	8.47%	8.52%	10.61%	8.43%	8.27%	8.47%	8.47%	8.52%	10.56%	8.44%	8.27%	8.47%	8.47%	8.52%	10.61%
70%	8.84%	7.23%	6.78%	6.78%	7.06%	7.55%	6.79%	6.66%	6.78%	6.78%	7.06%	7.53%	6.82%	6.66%	6.78%	6.78%	7.06%	7.55%
80%	7.12%	6.30%	6.55%	6.55%	6.47%	7.02%	6.32%	6.24%	6.55%	6.55%	6.47%	7.02%	6.33%	6.24%	6.55%	6.55%	6.47%	7.02%
90%	5.65%	5.79%	3.41%	3.41%	3.55%	4.87%	3.06%	2.95%	3.41%	3.41%	3.54%	4.87%	3.05%	2.95%	3.41%	3.41%	3.55%	4.87%
100%	0.34%	0.33%	0.34%	0.34%	0.36%	0.36%	0.29%	0.42%	0.26%	0.26%	0.31%	0.28%	0.29%	0.42%	0.26%	0.26%	0.31%	0.28%
Avg30	16.14%	15.53%	17.64%	17.64%	16.29%	16.04%	17.61%	17.48%	17.64%	17.64%	16.29%	15.91%	16.60%	17.48%	17.64%	17.64%	16.29%	16.04%
AVG100	10.34%	9.84%	10.09%	10.09%	9.60%	10.49%	10.03%	9.89%	10.08%	10.08%	9.59%	10.43%	9.73%	9.89%	10.08%	10.08%	9.59%	10.48%

T431 Query 8

T431 Q8	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	2.21%	2.21%	2.71%	3.50%	3.92%	2.23%	2.68%	2.23%	2.70%	6.36%	3.04%	2.22%	14.29%	2.22%	12.40%	12.40%	4.59%	2.25%
20%	0%	0%	0.51%	0.58%	4.18%	2.59%	1.15%	1.37%	1.15%	1.31%	3.71%	2.33%	2.93%	3.18%	2.95%	2.95%	3.76%	3.18%
30%	0%	0%	0.26%	0.27%	1.85%	0.36%	0.37%	0.34%	0.37%	0.40%	3.57%	0.36%	0.45%	0.33%	0.38%	0.38%	3.29%	0.33%
40%	0%	0%	0%	0%	0.50%	0.23%	0.21%	0.21%	0.21%	0.21%	1.16%	0.22%	0.24%	0.24%	0.23%	0.23%	1.96%	0.24%
50%	0%	0%	0%	0%	0.24%	0.21%	0.21%	0.19%	0.21%	0.21%	0.29%	0.20%	0.23%	0.22%	0.23%	0.23%	0.35%	0.22%
60%	0%	0%	0%	0%	0%	0%	0.14%	0.14%	0.14%	0.14%	0.20%	0.14%	0.14%	0.14%	0.14%	0.14%	0.21%	0.14%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.74%	0.74%	1.16%	1.45%	3.32%	1.73%	1.40%	1.31%	1.41%	2.69%	3.44%	1.63%	5.89%	1.91%	5.24%	5.24%	3.88%	1.92%
AVG100	0.22%	0.22%	0.35%	0.43%	1.07%	0.56%	0.47%	0.45%	0.48%	0.86%	1.20%	0.55%	1.83%	0.63%	1.63%	1.63%	1.42%	0.64%

T431 Query 9

T431 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.52%	0.57%	2.60%	29%	21.13%	35.71%	2.56%	3.31%	2.44%	28.85%	20.55%	34.88%	28.85%	30%	28.85%	28.85%	21.13%	30%
20%	0.35%	0.29%	0.72%	1.07%	9.23%	1.28%	0.66%	0.39%	0.75%	18.29%	11.54%	14.78%	16.39%	5.92%	18.18%	15.54%	12.66%	19.23%
30%	0.28%	0.23%	0.39%	0.52%	1.96%	0.60%	0.48%	0.32%	0.45%	0.94%	4.86%	0.77%	1.00%	0.94%	0.98%	0.98%	6.75%	0.93%
40%	0.22%	0.22%	0.24%	0.33%	0.65%	0.33%	0.28%	0.26%	0.26%	0.45%	1.03%	0.62%	0.55%	0.46%	0.48%	0.48%	1.13%	0.65%
50%	0.20%	0.19%	0.18%	0.20%	0.30%	0.13%	0.23%	0.21%	0.23%	0.25%	0.52%	0.22%	0.30%	0.27%	0.26%	0.26%	0.53%	0.25%
60%	0%	0%	0.15%	0.17%	0.19%	0.14%	0.17%	0.18%	0.17%	0.19%	0.29%	0.14%	0.19%	0.23%	0.20%	0.20%	0.34%	0.15%
70%	0%	0%	0%	0%	0%	0%	0.15%	0.16%	0.15%	0.15%	0.16%	0.13%	0.15%	0.17%	0.15%	0.15%	0.17%	0.13%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.14%	0.13%	0%	0%	0%	0%	0.14%	0.13%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.38%	0.36%	1.23%	10.34%	10.77%	12.53%	1.23%	1.34%	1.21%	16.03%	12.32%	16.81%	15.41%	12.29%	16.00%	15.12%	13.51%	16.72%
AVG100	0.16%	0.15%	0.43%	3.17%	3.35%	3.82%	0.45%	0.48%	0.44%	4.91%	3.91%	5.17%	4.74%	3.80%	4.91%	4.65%	4.28%	5.15%

T431 Query 10

T431 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.53%	1.35%	3.53%	11.28%	2.84%	4.73%	5.26%	3.64%	4.32%	10.14%	2.99%	4.73%	12%	10.87%	10.79%	10.95%	4.81%	6%
20%	0.90%	0.74%	0.96%	5.88%	1.40%	2.05%	1.13%	0.99%	1.10%	5.66%	1.53%	2.25%	7.65%	6.36%	6.62%	7.39%	2.36%	2.83%
30%	0.83%	0.77%	0.93%	0.93%	0.76%	0.96%	0.87%	0.86%	0.93%	0.92%	0.83%	1.19%	4.77%	4.57%	1.75%	4.90%	1.03%	1.72%
40%	0.59%	0.53%	0.57%	0.73%	0.43%	0.40%	0.67%	0.59%	0.66%	0.73%	0.46%	0.41%	0.81%	0.80%	0.91%	0.83%	0.62%	0.43%
50%	0.33%	0.35%	0.35%	0.37%	0.32%	0.38%	0.36%	0.35%	0.37%	0.37%	0.32%	0.38%	0.69%	0.73%	0.81%	0.74%	0.42%	0.44%
60%	0.33%	0.31%	0.32%	0.36%	0.29%	0.34%	0.36%	0.32%	0.33%	0.36%	0.29%	0.34%	0.51%	0.40%	0.39%	0.40%	0.33%	0.42%
70%	0%	0%	0%	0%	0.28%	0.28%	0%	0%	0%	0%	0.28%	0.28%	0.37%	0.39%	0.36%	0.39%	0.32%	0.35%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.23%	0.23%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.08%	0.96%	1.81%	6.03%	1.67%	2.58%	2.42%	1.83%	2.12%	5.57%	1.78%	2.72%	8.14%	7.27%	6.39%	7.75%	2.73%	3.52%
AVG100	0.45%	0.41%	0.67%	1.96%	0.63%	0.91%	0.86%	0.67%	0.77%	1.82%	0.67%	0.96%	2.68%	2.41%	2.16%	2.56%	1.01%	1.24%

T431 Query 1 entropy

T431 Q1	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	14.71%	16.48%	18.07%	18.07%	17.44%	17.86%	17.65%	17.44%	18.07%	18.07%	17.44%	17.86%	17.65%	17.44%	18.07%	18.07%	17.44%	17.86%
20%	19.11%	16.58%	19.11%	19.11%	18.52%	18.87%	20%	20.13%	19.11%	19.11%	18.52%	18.87%	20%	20.13%	19.11%	19.11%	18.52%	18.87%
30%	16.30%	18.18%	16.92%	16.92%	16.30%	16.67%	16.67%	16.92%	16.92%	16.92%	16.30%	16.67%	16.67%	16.92%	16.92%	16.92%	16.30%	16.67%
40%	14.57%	12.07%	13.29%	13.29%	13.29%	13.98%	13.26%	14.05%	13.29%	13.29%	13.29%	13.98%	13.26%	14.05%	13.29%	13.29%	13.29%	13.98%
50%	9.95%	10.38%	11.96%	11.96%	12.01%	12.56%	12.98%	11.84%	11.96%	11.96%	12.01%	12.44%	12.98%	11.84%	11.96%	11.96%	12.01%	12.56%
60%	9.96%	10.71%	8.75%	8.75%	9.35%	10.18%	9.75%	9.99%	8.75%	8.75%	9.33%	10.17%	9.81%	9.99%	8.75%	8.75%	9.35%	10.18%
70%	10.03%	9.95%	8.55%	8.55%	8.03%	7.80%	9.29%	8.38%	8.55%	8.55%	8.04%	7.77%	9.30%	8.38%	8.55%	8.55%	8.03%	7.80%
80%	9.73%	7.60%	4.68%	4.68%	4.85%	5.32%	4.53%	4.55%	4.68%	4.68%	4.85%	5.31%	4.54%	4.55%	4.68%	4.68%	4.85%	5.32%
90%	6.45%	5.04%	2.12%	2.12%	2.44%	2.74%	2.23%	2.12%	2.12%	2.12%	2.45%	2.74%	2.24%	2.12%	2.12%	2.12%	2.44%	2.74%
100%	0.36%	0.34%	0.52%	0.52%	0.51%	0.42%	0.62%	0.54%	0.46%	0.46%	0.50%	0.38%	0.62%	0.54%	0.46%	0.46%	0.50%	0.38%
Avg30	16.70%	17.08%	18.03%	18.03%	17.42%	17.80%	18.10%	18.17%	18.03%	18.03%	17.42%	17.80%	18.10%	18.17%	18.03%	18.03%	17.42%	17.80%
AVG100	11.11%	10.73%	10.40%	10.40%	10.27%	10.64%	10.70%	10.60%	10.39%	10.39%	10.27%	10.62%	10.71%	10.60%	10.39%	10.39%	10.27%	10.64%

T431 Query 2 entropy

T431 Q2	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	19.74%	20.27%	15.63%	25.42%	36.59%	18.52%	29.41%	31.25%	28.85%	25.42%	36.59%	18.52%	24.19%	29.41%	24.59%	25%	35.71%	17.86%
20%	23.08%	23.44%	20.27%	18.07%	21.13%	16.58%	17.24%	17.14%	17.24%	17.96%	21.13%	16.13%	17.86%	18.41%	17.86%	17.96%	20.27%	15.71%
30%	18.49%	18.80%	25.14%	14.67%	16.92%	15.33%	14.57%	14.43%	14.62%	14.57%	16.92%	15.07%	16%	15.23%	14.47%	14.57%	16.99%	14.33%
40%	16.95%	17.00%	21.00%	15.53%	18.44%	15.82%	12.69%	12.58%	12.83%	15.45%	18.44%	15.73%	15.73%	16.48%	15.29%	15.37%	18.27%	13.47%
50%	16.78%	16.78%	20.05%	12.87%	15.32%	12.50%	8.53%	8.30%	9.44%	12.67%	15.32%	12.31%	12.13%	13.31%	12.52%	12.59%	15.23%	10.34%
60%	16.15%	16.07%	17.73%	9.25%	12.09%	5.45%	8.08%	7.95%	8.45%	8.93%	12.09%	5.40%	9.26%	11.47%	8.90%	8.88%	12.54%	5.02%
70%	12.82%	12.22%	15.95%	8.65%	10.98%	5.56%	8.26%	8.18%	8.53%	8.43%	10.98%	5.39%	8%	8.90%	8.42%	8.25%	11.11%	5.05%
80%	9.87%	9.56%	13.69%	7.79%	9.42%	5.56%	7.67%	7.52%	7.80%	7.74%	9.42%	5.23%	7.46%	8.29%	7.75%	7.36%	9.66%	4.92%
90%	6.16%	5.42%	8.52%	6.20%	6.03%	4.73%	5.50%	5.60%	5.58%	6.16%	6.03%	4.05%	5.78%	6.40%	6.19%	5.85%	6.04%	3.84%
100%	4.37%	4.37%	3.79%	0.52%	2.39%	0.47%	0.28%	4.39%	0.27%	0.27%	1.69%	0.25%	0.31%	4.38%	0.27%	0.27%	1.62%	0.25%
Avg30	20.43%	20.84%	20.35%	19.39%	24.88%	16.81%	20.41%	20.94%	20.24%	19.32%	24.88%	16.57%	19.35%	21.01%	18.97%	19.18%	24.32%	15.97%
AVG100	14.44%	14.39%	16.18%	11.90%	14.93%	10.05%	11.22%	11.73%	11.36%	11.76%	14.86%	9.81%	11.67%	13.23%	11.63%	11.61%	14.74%	9.08%

T431 Query 3 entropy

T431 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	15.31%	15%	18.52%	18.52%	22.06%	15.63%	19.23%	20.55%	18.52%	18.52%	22.06%	15.63%	24.59%	21.43%	24.59%	24.59%	24.59%	21.43%
20%	20.55%	20.69%	22.06%	22.06%	22.90%	20.69%	24.79%	22.22%	22.06%	22.06%	22.90%	20.69%	16.58%	22.06%	16.85%	16.85%	16.85%	16.30%
30%	23.66%	24.04%	25.43%	25.43%	24.58%	18.88%	23.16%	22%	25.29%	25.29%	24.58%	18.88%	19.38%	19.30%	19.56%	19.56%	19.56%	15.88%
40%	23.70%	23.79%	25.21%	25.21%	24.58%	14.60%	20.07%	19.80%	25.11%	25.11%	24.58%	14.60%	17.83%	19.80%	17.30%	17.30%	21.22%	15.69%
50%	18.88%	18.55%	19.53%	19.53%	19.02%	6.57%	18.93%	17.83%	19.42%	19.42%	19.02%	6.51%	16.90%	20.27%	16.78%	16.78%	19.84%	14.15%
60%	17.91%	17.80%	17.94%	17.94%	17.80%	1.37%	17.49%	14.78%	17.80%	17.80%	17.80%	1.42%	16.51%	16.89%	16.04%	16.04%	17.80%	12.96%
70%	16.03%	15.81%	15.43%	15.43%	14.29%	1.42%	14.29%	11.03%	15.21%	15.21%	14.29%	1.42%	14.69%	14.59%	15.23%	15.23%	14.17%	11.05%
80%	13.79%	12.63%	13.07%	13.07%	10.19%	1.57%	12.33%	8.73%	12.69%	12.69%	10.19%	1.57%	10.91%	11.05%	10.85%	10.85%	11.01%	8.17%
90%	8.87%	8.36%	8.52%	8.52%	3.42%	1.70%	8.47%	7.28%	8.16%	8.16%	3.42%	1.70%	7.98%	7.30%	7.67%	7.67%	3.47%	5.50%
100%	3.01%	1.13%	2.49%	2.49%	1.03%	1.10%	2.05%	1.19%	2.32%	2.32%	1.05%	1.06%	1.87%	1.76%	1.70%	1.70%	1.05%	0.23%
Avg30	19.84%	19.91%	22.00%	22.00%	23.18%	18.40%	22.39%	21.59%	21.96%	21.96%	23.18%	18.40%	20.18%	20.93%	20.33%	20.33%	20.33%	17.87%
AVG100	16.17%	15.78%	16.82%	16.82%	15.99%	8.35%	16.08%	14.54%	16.66%	16.66%	15.99%	8.35%	14.72%	15.44%	14.66%	14.66%	14.96%	12.14%

T431 Query 4 entropy

T431 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	19.74%	24.59%	24.59%	24.59%	22.73%	20.55%	20.55%	24.59%	24.59%	24.59%	22.73%	20.55%	26.32%	14.29%	14.29%	14.29%	17.65%	31.25%
20%	12.45%	11.49%	11.49%	15%	13.83%	22.39%	15.31%	11.49%	11.49%	15%	13.83%	22.39%	14.85%	17.44%	17.44%	17.44%	20.98%	19.23%
30%	10.38%	11.58%	11.58%	14.19%	8.45%	15.44%	11.52%	11.58%	11.58%	14.19%	8.45%	15.44%	13.97%	18.72%	18.72%	18.72%	17.19%	15.71%
40%	11.24%	11.78%	11.78%	13.63%	8.49%	2.40%	12.42%	11.75%	11.75%	13.63%	8.48%	2.38%	8.98%	17.88%	17.88%	17.88%	18.21%	2.38%
50%	10.76%	11.16%	11.16%	12.59%	8.24%	2.74%	11.13%	11.11%	11.11%	12.56%	8.22%	2.72%	9.05%	15.98%	15.98%	15.98%	17.01%	2.72%
60%	8.67%	9.94%	9.94%	10.35%	7.84%	2.51%	8.65%	9.88%	9.88%	10.28%	7.82%	2.50%	9.88%	12.92%	12.92%	12.92%	13.89%	2.50%
70%	8.38%	9.46%	9.46%	9.94%	6.74%	2.88%	8.36%	9.41%	9.41%	9.88%	6.73%	2.86%	9.99%	10.83%	10.83%	10.83%	13.83%	2.87%
80%	7.99%	8.26%	8.26%	9.04%	5.16%	3.19%	8.09%	8.20%	8.20%	8.95%	5.15%	3.17%	8.06%	9.25%	9.25%	9.25%	10.64%	3.18%
90%	5.75%	5.92%	5.92%	5.99%	4.37%	3.49%	5.66%	5.88%	5.88%	5.95%	4.37%	3.46%	5.84%	6.20%	6.20%	6.20%	6.54%	3.47%
100%	2.91%	2.92%	2.92%	2.92%	0.42%	2.90%	2.89%	2.91%	2.91%	2.91%	0.42%	2.89%	2.91%	2.90%	2.90%	2.90%	2.90%	2.89%
Avg30	14.19%	15.89%	15.89%	17.93%	15.00%	19.46%	15.79%	15.89%	15.89%	17.93%	15.00%	19.46%	18.38%	16.82%	16.82%	16.82%	18.60%	22.07%
AVG100	9.83%	10.71%	10.71%	11.82%	8.62%	7.85%	10.46%	10.68%	10.68%	11.79%	8.62%	7.84%	10.98%	12.64%	12.64%	12.64%	13.88%	8.62%

T431 Query 5 entropy

T431 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	18.75%	19.48%	18.52%	18.75%	21.74%	24.19%	18.07%	15.79%	18.29%	18.52%	21.74%	24.19%	18.07%	15.79%	17.44%	17.44%	18.07%	24.59%
20%	21.28%	22.22%	20.83%	21.74%	21.43%	25.21%	18.29%	17.05%	20.41%	21.28%	21.43%	25.21%	21.13%	20.27%	20.98%	20.98%	20.83%	20.55%
30%	15.55%	21.05%	20.76%	20.37%	15.55%	16.42%	17.67%	15.39%	20.28%	20%	15.55%	16.42%	21.05%	18.33%	20.56%	20.56%	19.05%	14.82%
40%	16.34%	19.60%	21.07%	19.73%	15.99%	14.18%	18.15%	15.95%	20.56%	19.28%	15.99%	14.15%	17.83%	17.46%	19.60%	19.60%	17.93%	10.17%
50%	14.45%	17.49%	16.67%	16.63%	15.29%	6.54%	13.81%	13.17%	16.19%	16.16%	15.29%	6.49%	16.63%	14.77%	16.34%	16.34%	16.05%	7.69%
60%	12.55%	15.03%	14.22%	14.29%	12.90%	2.73%	12.62%	11.71%	13.82%	13.89%	12.90%	2.72%	12.48%	12.36%	13.89%	13.89%	13.22%	2.80%
70%	11.54%	13.42%	12.89%	13.37%	12.84%	3.11%	10.92%	10.96%	12.50%	12.95%	12.84%	3.10%	11.93%	11.69%	13.05%	13.05%	13.32%	3.13%
80%	10.56%	10.72%	10.92%	11.20%	10.38%	3.37%	9.64%	10.25%	10.58%	10.85%	10.38%	3.36%	10.04%	10.92%	11.06%	11.06%	10.36%	3.42%
90%	6.66%	6.72%	6.19%	6.79%	6.48%	3.66%	5.46%	6.47%	5.85%	6.40%	6.48%	3.64%	6.09%	7.17%	6.61%	6.61%	6.46%	3.67%
100%	3.06%	3.08%	2.34%	2.34%	2.23%	2.22%	1.76%	3.06%	1.76%	1.76%	1.77%	1.75%	1.77%	3.07%	1.76%	1.76%	1.96%	0.78%
Avg30	18.53%	20.92%	20.04%	20.29%	19.57%	21.94%	18.01%	16.07%	19.66%	19.93%	19.57%	21.94%	20.08%	18.13%	19.66%	19.66%	19.32%	19.98%
AVG100	13.07%	14.88%	14.44%	14.52%	13.48%	10.16%	12.64%	11.98%	14.02%	14.11%	13.44%	10.10%	13.70%	13.18%	14.13%	14.13%	13.73%	9.16%

T431 Query 6 entropy

T431 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0%	0%	1.62%	1.62%	14.56%	7.46%	3.93%	8.57%	3.88%	3.88%	14.29%	26.79%	53.57%	26.79%	51.72%	51.72%	44.12%	31.92%
20%	0%	0%	0.96%	0.96%	0.69%	0.95%	1.84%	0.83%	1.88%	1.88%	2.75%	0.83%	25.86%	27.78%	27.52%	27.52%	28.04%	28.57%
30%	0%	0%	0.36%	0.36%	0.44%	0.27%	1.21%	0.56%	1.21%	1.21%	1.96%	0.56%	3.63%	1.95%	1.13%	1.13%	2.84%	1.55%
40%	0%	0%	0%	0%	0.20%	0.16%	0.72%	0.37%	0.68%	0.68%	1.11%	0.37%	0.94%	0.48%	0.47%	0.47%	0.69%	0.49%
50%	0%	0%	0%	0%	0%	0%	0.46%	0.27%	0.47%	0.47%	0.61%	0.27%	0.42%	0.31%	0.27%	0.27%	0.43%	0.32%
60%	0%	0%	0%	0%	0%	0%	0.31%	0.22%	0.29%	0.29%	0.32%	0.22%	0.26%	0.23%	0.22%	0.22%	0.30%	0.23%
70%	0%	0%	0%	0%	0%	0%	0.18%	0.18%	0.18%	0.18%	0.21%	0.18%	0.18%	0.18%	0.18%	0.18%	0.19%	0.18%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.00%	0.00%	0.98%	0.98%	5.23%	2.89%	2.33%	3.32%	2.32%	2.32%	6.33%	9.39%	27.69%	18.84%	26.79%	26.79%	25.00%	20.68%
AVG100	0.00%	0.00%	0.29%	0.29%	1.59%	0.88%	0.86%	1.10%	0.86%	0.86%	2.12%	2.92%	8.49%	5.77%	8.15%	8.15%	7.66%	6.33%

T431 Query 7 entropy

T431 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0%	0%	5.28%	5.28%	23.44%	15.46%	5.68%	14.42%	5.70%	5.70%	19.48%	14.42%	8.93%	15.96%	21.74%	21.74%	21.74%	15.96%
20%	0%	0%	0.99%	0.99%	0.76%	0.91%	2.05%	1.29%	2.01%	2.01%	4.62%	1.29%	2.46%	1.87%	2.50%	2.50%	5.39%	1.87%
30%	0%	0%	0.48%	0.48%	0.61%	0.30%	1.22%	0.74%	1.20%	1.20%	1.18%	0.74%	1.70%	1.21%	1.74%	1.74%	0.93%	1.21%
40%	0%	0%	0.20%	0.20%	0.25%	0.17%	0.78%	0.44%	0.79%	0.79%	1.09%	0.44%	1.03%	0.79%	1.03%	1.03%	0.77%	0.79%
50%	0%	0%	0%	0%	0%	0%	0.50%	0.34%	0.47%	0.47%	0.86%	0.34%	0.49%	0.52%	0.48%	0.48%	0.63%	0.52%
60%	0%	0%	0%	0%	0%	0%	0.25%	0.23%	0.25%	0.25%	0.45%	0.23%	0.38%	0.38%	0.35%	0.35%	0.43%	0.38%
70%	0%	0%	0%	0%	0%	0%	0.18%	0.18%	0.18%	0.18%	0.27%	0.18%	0.19%	0.21%	0.19%	0.19%	0.30%	0.21%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.00%	0.00%	2.25%	2.25%	8.27%	5.56%	2.98%	5.48%	2.97%	2.97%	8.43%	5.48%	4.36%	6.34%	8.66%	8.66%	9.35%	6.34%
AVG100	0.00%	0.00%	0.69%	0.69%	2.51%	1.68%	1.07%	1.76%	1.06%	1.06%	2.79%	1.76%	1.52%	2.09%	2.80%	2.80%	3.02%	2.09%

T431 Query 8 entropy

T431 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	2.21%	2.21%	2.50%	2.96%	4.03%	2.28%	2.42%	2.50%	2.46%	9.80%	3.93%	2.23%	2.22%	2.22%	17.44%	17.05%	5.36%	2.34%
20%	0%	0%	0.47%	0.58%	4.27%	2.24%	0.98%	1.05%	0.99%	1.11%	3.80%	1.83%	2.85%	2.85%	2.34%	1.82%	3.97%	2.85%
30%	0%	0%	0.21%	0.22%	2.10%	0.35%	0.44%	0.37%	0.40%	0.44%	3.13%	0.35%	0.48%	0.48%	0.36%	0.55%	3.38%	0.48%
40%	0%	0%	0%	0%	0.52%	0.23%	0.26%	0.26%	0.26%	0.26%	0.72%	0.22%	0.29%	0.24%	0.22%	0.28%	1.71%	0.24%
50%	0%	0%	0%	0%	0.24%	0.21%	0.17%	0.19%	0.18%	0.18%	0.32%	0.21%	0.20%	0.22%	0.22%	0.21%	0.33%	0.22%
60%	0%	0%	0%	0%	0%	0%	0.14%	0.14%	0.14%	0.14%	0.18%	0.14%	0.14%	0.14%	0.14%	0.14%	0.21%	0.14%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.74%	0.74%	1.06%	1.25%	3.47%	1.62%	1.28%	1.31%	1.28%	3.78%	3.62%	1.47%	1.85%	1.85%	6.71%	6.47%	4.24%	1.89%
AVG100	0.22%	0.22%	0.32%	0.38%	1.12%	0.53%	0.44%	0.45%	0.44%	1.19%	1.21%	0.50%	0.62%	0.61%	2.07%	2.00%	1.50%	0.63%

T431 Query 9 entropy

T431 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.53%	0.55%	3.46%	30%	20.27%	31.92%	3.43%	3.75%	3.45%	26.32%	20.55%	27.78%	25%	30%	24.59%	24.59%	22.73%	27.27%
20%	0.33%	0.35%	0.39%	1.05%	9.26%	1.92%	0.38%	0.39%	0.38%	14.35%	9.87%	5.59%	13.04%	2.53%	13.64%	13.10%	11.91%	13.89%
30%	0.28%	0.27%	0.34%	0.71%	2.68%	0.54%	0.34%	0.35%	0.35%	0.79%	4.83%	1.11%	1.18%	0.76%	1.21%	1.21%	2.53%	1.00%
40%	0.23%	0.21%	0.29%	0.32%	0.53%	0.37%	0.29%	0.29%	0.29%	0.53%	1.66%	0.48%	0.52%	0.62%	0.55%	0.55%	1.14%	0.56%
50%	0.21%	0.19%	0.22%	0.21%	0.34%	0.15%	0.25%	0.19%	0.26%	0.28%	0.40%	0.30%	0.36%	0.27%	0.33%	0.33%	0.51%	0.31%
60%	0%	0%	0.18%	0.16%	0.20%	0.15%	0.19%	0.18%	0.19%	0.21%	0.30%	0.13%	0.22%	0.22%	0.23%	0.23%	0.29%	0.14%
70%	0%	0%	0%	0%	0%	0%	0.17%	0.17%	0.17%	0.16%	0.18%	0.13%	0.17%	0.17%	0.16%	0.16%	0.19%	0.13%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.14%	0.13%	0%	0%	0%	0%	0.15%	0.13%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.38%	0.39%	1.40%	10.59%	10.74%	11.46%	1.38%	1.49%	1.39%	13.82%	11.75%	11.49%	13.08%	11.10%	13.15%	12.97%	12.39%	14.05%
AVG100	0.16%	0.16%	0.49%	3.25%	3.33%	3.50%	0.50%	0.53%	0.51%	4.26%	3.79%	3.56%	4.05%	3.46%	4.07%	4.02%	3.94%	4.34%

T431 Query 10 entropy

T431 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	2.15%	1.77%	3.79%	11.63%	3.25%	4.57%	4.93%	3.70%	4.55%	11.45%	3.23%	4.59%	10.95%	10.95%	11.28%	11.11%	4.67%	5.16%
20%	0.82%	0.85%	1.14%	3.43%	1.65%	1.72%	1.25%	0.93%	1.16%	3.37%	1.74%	1.70%	4.33%	4.08%	3.50%	4.47%	2.77%	2.51%
30%	0.74%	0.77%	0.77%	1.18%	0.76%	1.02%	0.75%	0.84%	0.81%	1.16%	0.77%	1.31%	3.40%	3.47%	1.24%	3.49%	1.43%	1.93%
40%	0.56%	0.57%	0.59%	0.71%	0.72%	0.45%	0.62%	0.60%	0.66%	0.71%	0.79%	0.46%	0.78%	0.91%	0.77%	0.81%	0.78%	0.48%
50%	0.31%	0.32%	0.33%	0.49%	0.34%	0.38%	0.45%	0.34%	0.44%	0.49%	0.34%	0.39%	0.68%	0.70%	0.62%	0.69%	0.59%	0.43%
60%	0.31%	0.32%	0.32%	0.33%	0.28%	0.35%	0.33%	0.32%	0.34%	0.33%	0.28%	0.36%	0.51%	0.49%	0.38%	0.50%	0.30%	0.43%
70%	0%	0%	0%	0%	0.27%	0.28%	0%	0%	0%	0%	0.27%	0.28%	0.35%	0.36%	0.34%	0.36%	0.31%	0.36%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.22%	0.23%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.24%	1.13%	1.90%	5.41%	1.89%	2.44%	2.31%	1.82%	2.17%	5.33%	1.91%	2.53%	6.23%	6.17%	5.34%	6.36%	2.96%	3.20%
AVG100	0.49%	0.46%	0.69%	1.78%	0.73%	0.88%	0.83%	0.67%	0.80%	1.75%	0.74%	0.91%	2.10%	2.10%	1.81%	2.14%	1.11%	1.15%

T431 Query 1 idf

T431 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	14.56%	14.71%	19.48%	19.48%	15.79%	16.48%	18.99%	20.00%	19.48%	19.48%	15.79%	16.48%	14.56%	20%	19.48%	19.48%	15.79%	16.48%
20%	16.39%	15.39%	16.95%	16.95%	18.18%	16.95%	17%	16.76%	16.95%	16.95%	18.18%	16.58%	16.39%	16.76%	16.95%	16.95%	18.18%	16.95%
30%	18.64%	17.74%	16.60%	16.60%	16.06%	14.87%	17.12%	16.18%	16.60%	16.60%	16.06%	14.87%	18.64%	16.18%	16.60%	16.60%	16.06%	14.87%
40%	10.81%	10.95%	12.45%	12.45%	14.08%	14.05%	12.34%	13.05%	12.45%	12.45%	14.08%	14.05%	10.81%	13.05%	12.45%	12.45%	14.08%	14.05%
50%	10.62%	11.39%	10.19%	10.19%	11.64%	12.40%	10.84%	10.00%	10.19%	10.19%	11.64%	12.31%	10.62%	10%	10.19%	10.19%	11.64%	12.40%
60%	11.47%	11.74%	8.46%	8.46%	11.44%	10.70%	8.38%	8.30%	8.46%	8.46%	11.44%	10.60%	11.47%	8.30%	8.46%	8.46%	11.44%	10.70%
70%	9.04%	8.24%	6.78%	6.78%	8.35%	7.61%	6.78%	6.65%	6.78%	6.78%	8.35%	7.59%	9.04%	6.65%	6.78%	6.78%	8.35%	7.61%
80%	7.17%	7.27%	6.53%	6.53%	7.57%	6.92%	6.18%	6.19%	6.53%	6.53%	7.57%	6.92%	7.17%	6.19%	6.53%	6.53%	7.57%	6.92%
90%	5.61%	5.72%	3.26%	3.26%	5.13%	4.77%	3.05%	2.84%	3.26%	3.26%	5.13%	4.76%	5.61%	2.84%	3.26%	3.26%	5.13%	4.77%
100%	0.34%	0.33%	0.35%	0.35%	0.37%	0.36%	0.30%	0.43%	0.27%	0.27%	0.34%	0.29%	0.34%	0.43%	0.27%	0.27%	0.34%	0.29%
Avg30	16.53%	15.94%	17.68%	17.68%	16.68%	16.10%	17.72%	17.65%	17.68%	17.68%	16.68%	15.97%	16.53%	17.65%	17.68%	17.68%	16.68%	16.10%
AVG100	10.46%	10.35%	10.11%	10.11%	10.86%	10.51%	10.10%	10.04%	10.10%	10.10%	10.86%	10.44%	10.46%	10.04%	10.10%	10.10%	10.86%	10.50%

T431 Query 2 idf

T431 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	19.48%	20.27%	16.13%	22.73%	32.61%	17.44%	18.07%	28.30%	16.13%	22.73%	32.61%	17.44%	19.48%	25%	22.39%	22.73%	33.33%	17.44%
20%	23.08%	23.08%	20.98%	17.44%	19.23%	16.76%	20.98%	15.87%	20.98%	17.24%	19.23%	16.30%	23.08%	19.48%	17.24%	17.14%	19.23%	16.22%
30%	18.64%	18.64%	23.91%	15.88%	19.56%	14.87%	23.28%	14.29%	23.91%	15.77%	19.56%	15.33%	18.64%	15.39%	14.92%	15.60%	19.73%	13.84%
40%	16.91%	16.30%	20.07%	13.66%	17.56%	13.35%	20.56%	12.40%	20.07%	13.50%	17.56%	13.26%	16.91%	12.58%	12.53%	13.35%	17.88%	11.52%
50%	16.74%	16.30%	20.22%	12.44%	15.35%	10.95%	20.22%	10.11%	20.22%	12.25%	15.35%	11.13%	16.74%	9.46%	10.56%	11.94%	15.55%	8.19%
60%	15.95%	16.18%	17.98%	9.85%	11.53%	6.04%	17.59%	8.44%	17.98%	9.80%	11.53%	6.03%	15.95%	8.82%	9.29%	9.34%	11.74%	5.77%
70%	12.34%	12.02%	15.12%	9.30%	11.20%	6.23%	14.92%	8.82%	15.09%	9.34%	11.22%	6.21%	12.34%	8.52%	8.75%	8.68%	11.26%	5.90%
80%	9.54%	9.90%	12.22%	8.36%	8.85%	6.15%	12.13%	7.41%	12.17%	8.36%	8.85%	6.10%	9.54%	7.85%	8.00%	7.86%	8.89%	5.73%
90%	5.45%	5.30%	9.02%	5.91%	5.88%	4.94%	9.32%	5.82%	8.97%	5.90%	5.88%	4.71%	5.45%	5.63%	5.48%	5.60%	5.77%	4.19%
100%	4.37%	4.36%	3.59%	0.47%	0.53%	0.46%	3.78%	4.38%	3.02%	0.25%	0.31%	0.25%	4.37%	4.14%	0.25%	0.25%	0.31%	0.25%
Avg30	20.40%	20.66%	20.34%	18.68%	23.80%	16.36%	20.78%	19.49%	20.34%	18.58%	23.80%	16.36%	20.40%	19.96%	18.18%	18.49%	24.10%	15.83%
AVG100	14.25%	14.24%	15.92%	11.60%	14.23%	9.72%	16.08%	11.58%	15.85%	11.51%	14.21%	9.68%	14.25%	11.69%	10.94%	11.25%	14.37%	8.90%

T431 Query 3 idf

T431 Q3	**ko+** **ENTIRE**	**ko+** **BASE**	**ko+** **SUB**	**ko+** **SUPER**	**ko+** **SEP**	**ko+** **SEP_FC**	**sub+** **ENTIRE**	**sub+** **BASE**	**sub+** **SUB**	**sub+** **SUPER**	**sub+** **SEP**	**sub+** **SEP_FC**	**super+** **ENTIRE**	**super+** **BASE**	**super+** **SUB**	**super+** **SUPER**	**super+** **SEP**	**super+** **SEP_FC**
10%	15.15%	18%	18.07%	18.07%	22.39%	15.31%	17.86%	19.23%	18.07%	18.07%	22.39%	15.31%	15.15%	21.43%	26.32%	26.32%	21.13%	19.48%
20%	20.69%	20.27%	23.44%	23.44%	23.81%	21.74%	22.73%	21.58%	23.44%	23.44%	23.81%	21.74%	20.69%	19.23%	16.30%	16.30%	19.74%	17.14%
30%	23.91%	23.53%	23.16%	23.16%	24.31%	19.05%	25.29%	23%	23.16%	23.16%	24.31%	19.05%	23.91%	20.18%	18.33%	18.33%	19.91%	15.28%
40%	24.38%	20.07%	23.79%	23.79%	22.52%	14.64%	24.58%	24.18%	23.79%	23.79%	22.52%	14.64%	24.38%	20.07%	17.25%	17.25%	21.22%	13.08%
50%	18.50%	18.09%	19.63%	19.63%	19.68%	7.70%	19.68%	19.79%	19.63%	19.63%	19.68%	10.11%	18.50%	19.22%	17.70%	17.70%	19.73%	12.15%
60%	17.84%	17.22%	17.66%	17.66%	17.87%	6.61%	18.28%	17.66%	17.66%	17.66%	17.87%	6.62%	17.84%	17.73%	16.76%	16.76%	17.80%	9.12%
70%	15.85%	14.57%	15.03%	15.03%	14.35%	1.56%	15.76%	14.33%	15.03%	15.03%	14.35%	1.57%	15.85%	16.56%	14.31%	14.31%	14.51%	9.16%
80%	14.12%	11.66%	13.13%	13.13%	10.42%	1.60%	13.20%	12.98%	13.11%	13.11%	10.42%	1.60%	14.12%	12.98%	10.51%	10.51%	11.07%	7.01%
90%	9.22%	7.75%	9.65%	9.65%	5.33%	1.68%	9.93%	8.05%	9.62%	9.62%	5.33%	1.68%	9.22%	9.10%	7.20%	7.20%	5.36%	2.81%
100%	3.00%	1.13%	3.00%	3.00%	1.03%	1.16%	3.22%	1.16%	2.94%	2.94%	1.06%	1.14%	3.00%	1.16%	1.27%	1.27%	1.06%	0.23%
Avg30	19.92%	20.55%	21.56%	21.56%	23.50%	18.70%	21.96%	21.41%	21.56%	21.56%	23.50%	18.70%	19.92%	20.28%	20.32%	20.32%	20.26%	17.30%
AVG100	16.27%	15.21%	16.65%	16.65%	16.17%	9.10%	17.05%	16.24%	16.64%	16.64%	16.17%	9.34%	16.27%	15.77%	14.59%	14.59%	15.15%	10.55%

T431 Query 4 idf

T431 Q4	**ko+** **ENTIRE**	**ko+** **BASE**	**ko+** **SUB**	**ko+** **SUPER**	**ko+** **SEP**	**ko+** **SEP_FC**	**sub+** **ENTIRE**	**sub+** **BASE**	**sub+** **SUB**	**sub+** **SUPER**	**sub+** **SEP**	**sub+** **SEP_FC**	**super+** **ENTIRE**	**super+** **BASE**	**super+** **SUB**	**super+** **SUPER**	**super+** **SEP**	**super+** **SEP_FC**
10%	22.39%	14.02%	13.89%	30.00%	18.52%	16.67%	15.79%	13.89%	13.89%	30%	18.52%	16.67%	22.39%	22.39%	22.39%	22.39%	16.67%	29.41%
20%	9.17%	12.05%	12.05%	14%	12.61%	20.98%	14.42%	11.95%	11.95%	14.22%	12.61%	20.98%	9.17%	14.93%	14.93%	14.93%	18.63%	17.05%
30%	7.69%	14.47%	14.47%	8.91%	6.74%	15.28%	10.23%	14.38%	14.38%	8.91%	6.72%	15.28%	7.69%	12.98%	12.98%	12.98%	16.24%	15.33%
40%	7.76%	10.16%	10.23%	9.16%	6.50%	2.40%	11.24%	9.98%	9.98%	9.11%	6.48%	2.38%	7.76%	10.91%	10.91%	10.91%	16.21%	2.38%
50%	7.68%	10.41%	10.45%	8.80%	6.46%	2.72%	8.58%	10.22%	10.22%	8.74%	6.45%	2.70%	7.68%	9.24%	9.24%	9.24%	14.83%	2.70%
60%	7.89%	9.59%	9.59%	9.13%	6.27%	2.51%	8.42%	9.42%	9.42%	9.07%	6.26%	2.50%	7.89%	9.46%	9.46%	9.46%	12.55%	2.50%
70%	8.21%	9.10%	9.10%	8.16%	5.81%	2.88%	8.21%	8.97%	8.97%	8.12%	5.81%	2.86%	8.21%	9.86%	9.86%	9.86%	13.20%	2.87%
80%	8.03%	8.47%	8.47%	6.67%	4.78%	3.19%	7.79%	8.37%	8.37%	6.64%	4.77%	3.17%	8.03%	8.17%	8.17%	8.17%	10.83%	3.18%
90%	6.02%	5.46%	5.46%	4.61%	4.16%	3.49%	5.70%	5.42%	5.42%	4.60%	4.16%	3.46%	6.02%	4.74%	4.74%	4.74%	6.18%	3.47%
100%	2.91%	2.92%	2.92%	2.92%	0.42%	2.90%	2.89%	2.91%	2.91%	2.91%	0.42%	2.89%	2.91%	2.91%	2.91%	2.91%	2.90%	2.89%
Avg30	13.08%	13.51%	13.47%	17.71%	12.62%	17.64%	13.48%	13.41%	13.41%	17.71%	12.61%	17.64%	13.08%	16.76%	16.76%	16.76%	17.18%	20.60%
AVG100	8.78%	9.66%	9.66%	10.26%	7.23%	7.30%	9.33%	9.55%	9.55%	10.23%	7.22%	7.29%	8.78%	10.56%	10.56%	10.56%	12.82%	8.18%

T431 Query 5 idf

T431 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	18.52%	19.48%	17.24%	18.75%	21.13%	25%	18.29%	16.13%	17.65%	18.52%	21.13%	25%	18.52%	15.79%	17.65%	17.65%	17.86%	20.27%
20%	18.99%	21.13%	18.29%	21.13%	17.96%	22.06%	18.52%	15.71%	18.18%	20.69%	17.96%	22.06%	18.99%	20%	20.55%	20.55%	19.61%	19.48%
30%	17.96%	20.56%	21.15%	20.18%	14.77%	19.73%	17.81%	13.33%	19.82%	19.82%	14.77%	20.09%	17.96%	15.94%	19.73%	19.73%	15.28%	20.66%
40%	18.21%	18.55%	18.79%	19.03%	14.64%	11.92%	18.32%	14.15%	18.55%	18.44%	14.64%	11.82%	18.21%	14.53%	18.85%	18.85%	15.78%	11.13%
50%	15.26%	17.25%	16.78%	16.05%	13.36%	5.20%	14.26%	12.40%	16.34%	15.65%	13.36%	5.18%	15.26%	13.50%	15.61%	15.61%	14.04%	5.24%
60%	13.71%	14.45%	14.29%	14.08%	12.14%	2.72%	12.19%	11.17%	13.91%	13.69%	12.14%	2.72%	13.71%	11.56%	13.74%	13.74%	12.24%	2.71%
70%	11.95%	12.29%	12.38%	12.64%	12.21%	3.09%	10.80%	10.33%	11.97%	12.25%	12.21%	3.08%	11.95%	10.83%	12.37%	12.37%	12.38%	3.08%
80%	10.91%	10.81%	9.82%	10.23%	10.07%	3.37%	9.44%	9.39%	9.52%	9.90%	10.07%	3.36%	10.91%	9.70%	10.03%	10.03%	10.37%	3.36%
90%	6.47%	6.29%	6.22%	6.46%	6.10%	3.64%	5.45%	5.91%	5.14%	6.15%	6.10%	3.62%	6.47%	6.41%	6.17%	6.17%	6.13%	3.59%
100%	3.06%	3.09%	2.34%	2.34%	2.23%	1.16%	1.76%	3.07%	1.76%	1.76%	1.77%	1.02%	3.06%	3.07%	1.76%	1.76%	1.77%	1.75%
Avg30	18.49%	20.39%	18.90%	20.02%	17.95%	22.26%	18.21%	15.06%	18.55%	19.68%	17.95%	22.38%	18.49%	17.24%	19.31%	19.31%	17.58%	20.14%
AVG100	13.50%	14.39%	13.73%	14.09%	12.46%	9.79%	12.68%	11.16%	13.28%	13.69%	12.41%	9.79%	13.50%	12.13%	13.64%	13.64%	12.54%	9.13%

T431 Query 6 idf

T431 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0%	0%	1.96%	1.96%	11.03%	7.35%	3.96%	10.64%	3.97%	3.97%	13.64%	10.71%	25.86%	25.86%	48.39%	48.39%	39.47%	31.25%
20%	0%	0%	1.15%	1.15%	0.66%	0.95%	1.95%	0.63%	1.93%	1.93%	4.14%	0.63%	30.93%	30.93%	31.25%	31.25%	29.41%	31.25%
30%	0%	0%	0.29%	0.29%	0.44%	0.26%	1.22%	0.45%	1.19%	1.19%	1.18%	0.45%	0.84%	0.84%	0.84%	0.84%	4.44%	0.84%
40%	0%	0%	0%	0%	0.19%	0.16%	0.66%	0.35%	0.71%	0.71%	1.02%	0.35%	0.41%	0.41%	0.41%	0.41%	0.90%	0.41%
50%	0%	0%	0%	0%	0%	0%	0.33%	0.26%	0.34%	0.34%	0.62%	0.26%	0.28%	0.28%	0.28%	0.28%	0.43%	0.28%
60%	0%	0%	0%	0%	0%	0%	0.25%	0.22%	0.26%	0.26%	0.29%	0.21%	0.22%	0.22%	0.22%	0.22%	0.28%	0.22%
70%	0%	0%	0%	0%	0%	0%	0.18%	0.18%	0.19%	0.19%	0.20%	0.19%	0.18%	0.18%	0.18%	0.18%	0.20%	0.19%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.00%	0.00%	1.14%	1.14%	4.04%	2.85%	2.38%	3.90%	2.36%	2.36%	6.32%	3.93%	19.21%	19.21%	26.83%	26.83%	24.44%	21.11%
AVG100	0.00%	0.00%	0.34%	0.34%	1.23%	0.87%	0.85%	1.27%	0.86%	0.86%	2.11%	1.28%	5.87%	5.87%	8.16%	8.16%	7.51%	6.44%

T431 Query 7 idf

T431 Q7	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0%	0%	6.20%	6.20%	25.00%	15.15%	6.47%	16.67%	6.41%	6.41%	21.13%	16.67%	21.13%	13.39%	8.52%	8.52%	19.48%	13.39%
20%	0%	0%	1.05%	1.05%	0.79%	0.57%	2.13%	0.72%	2.14%	2.14%	1.56%	0.72%	1.56%	2.32%	2.32%	2.32%	4.01%	2.32%
30%	0%	0%	0.36%	0.36%	0.72%	0.29%	1.19%	0.49%	1.18%	1.18%	1.10%	0.49%	1.10%	1.62%	1.62%	1.62%	1.09%	1.62%
40%	0%	0%	0.19%	0.19%	0.23%	0.17%	0.75%	0.39%	0.76%	0.76%	1.02%	0.39%	1.11%	1.11%	1.11%	1.11%	1.01%	1.11%
50%	0%	0%	0%	0%	0%	0%	0.35%	0.30%	0.35%	0.35%	0.77%	0.30%	0.44%	0.44%	0.44%	0.44%	0.74%	0.44%
60%	0%	0%	0%	0%	0%	0%	0.26%	0.22%	0.25%	0.25%	0.35%	0.22%	0.29%	0.29%	0.29%	0.29%	0.33%	0.29%
70%	0%	0%	0%	0%	0%	0%	0.18%	0.18%	0.18%	0.18%	0.23%	0.18%	0.19%	0.19%	0.19%	0.19%	0.22%	0.19%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.00%	0.00%	2.54%	2.54%	8.84%	5.34%	3.26%	5.96%	3.24%	3.24%	7.93%	5.96%	7.93%	5.78%	4.16%	4.16%	8.19%	5.78%
AVG100	0.00%	0.00%	0.78%	0.78%	2.67%	1.62%	1.13%	1.90%	1.13%	1.13%	2.61%	1.90%	2.58%	1.94%	1.45%	1.45%	2.69%	1.94%

T431 Query 8 idf

T431 Q8	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	2.21%	2.21%	2.74%	3.45%	3.92%	2.23%	2.65%	2.22%	2.72%	6.55%	3.10%	2.22%	12.93%	2.22%	12.93%	12.93%	4.64%	2.25%
20%	0%	0%	0.51%	0.56%	4.25%	2.52%	1.06%	1.34%	1.15%	1.30%	3.73%	2.20%	2.94%	3.00%	2.94%	2.94%	3.81%	3.09%
30%	0%	0%	0.26%	0.26%	1.84%	0.36%	0.36%	0.31%	0.39%	0.40%	3.49%	0.37%	0.39%	0.45%	0.39%	0.39%	3.24%	0.35%
40%	0%	0%	0%	0%	0.50%	0.23%	0.22%	0.21%	0.20%	0.21%	1.17%	0.22%	0.23%	0.24%	0.23%	0.23%	2.11%	0.24%
50%	0%	0%	0%	0%	0.24%	0.21%	0.21%	0.20%	0.21%	0.22%	0.29%	0.20%	0.23%	0.23%	0.23%	0.23%	0.35%	0.22%
60%	0%	0%	0%	0%	0%	0%	0.14%	0.14%	0.14%	0.14%	0.20%	0.14%	0.14%	0.14%	0.14%	0.14%	0.21%	0.14%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.74%	0.74%	1.17%	1.42%	3.34%	1.70%	1.36%	1.29%	1.42%	2.75%	3.44%	1.59%	5.42%	1.89%	5.42%	5.42%	3.90%	1.90%
AVG100	0.22%	0.22%	0.35%	0.43%	1.07%	0.55%	0.46%	0.44%	0.48%	0.88%	1.20%	0.53%	1.69%	0.63%	1.69%	1.69%	1.44%	0.63%

T431 Query 9 idf

T431 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.50%	0.56%	2.51%	31%	21.13%	36.59%	2.48%	3.56%	2.51%	28.85%	20.55%	34.88%	34.88%	30%	30%	30%	21.13%	28.30%
20%	0.35%	0.31%	0.69%	1.06%	9.01%	1.28%	0.68%	0.37%	0.69%	18.29%	11.63%	14.02%	14.02%	5.15%	17.96%	15.96%	12.35%	19.11%
30%	0.28%	0.24%	0.39%	0.52%	2.00%	0.62%	0.49%	0.33%	0.45%	0.89%	4.73%	0.75%	0.75%	0.91%	0.96%	0.96%	6.46%	0.97%
40%	0.22%	0.22%	0.25%	0.33%	0.66%	0.34%	0.28%	0.26%	0.25%	0.46%	1.04%	0.60%	0.46%	0.46%	0.50%	0.50%	1.16%	0.68%
50%	0.21%	0.20%	0.19%	0.20%	0.30%	0.14%	0.23%	0.21%	0.23%	0.25%	0.52%	0.24%	0.28%	0.28%	0.29%	0.27%	0.50%	0.26%
60%	0%	0%	0.15%	0.17%	0.19%	0.15%	0.17%	0.18%	0.17%	0.19%	0.28%	0.14%	0.23%	0.23%	0.20%	0.20%	0.34%	0.14%
70%	0%	0%	0%	0%	0%	0%	0.15%	0.17%	0.15%	0.15%	0.16%	0.13%	0.17%	0.17%	0.15%	0.15%	0.17%	0.13%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.14%	0.13%	0%	0%	0%	0%	0.14%	0.13%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.38%	0.37%	1.20%	10.94%	10.71%	12.83%	1.22%	1.42%	1.22%	16.01%	12.30%	16.55%	16.55%	12.02%	16.31%	15.64%	13.31%	16.13%
AVG100	0.16%	0.15%	0.42%	3.35%	3.33%	3.91%	0.45%	0.51%	0.45%	4.91%	3.90%	5.09%	5.08%	3.72%	5.01%	4.80%	4.22%	4.97%

T431 Query 10 idf

T431 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	1.52%	1.36%	3.51%	12.10%	3.02%	4.69%	5.36%	3.66%	4.67%	10.64%	3.39%	4.69%	4.35%	10.79%	11.54%	10.87%	4.35%	5.95%
20%	0.90%	0.73%	0.97%	5.92%	1.48%	2.04%	1.12%	0.99%	1.10%	5.64%	1.46%	2.17%	2.50%	6.33%	6.88%	7.39%	2.50%	2.83%
30%	0.82%	0.75%	0.88%	0.96%	0.76%	0.96%	0.86%	0.88%	0.93%	0.95%	0.88%	1.21%	1.01%	4.55%	1.08%	4.84%	1.01%	1.72%
40%	0.59%	0.52%	0.57%	0.72%	0.46%	0.41%	0.72%	0.57%	0.63%	0.71%	0.56%	0.41%	0.51%	0.81%	0.81%	0.81%	0.51%	0.49%
50%	0.33%	0.34%	0.37%	0.51%	0.32%	0.38%	0.36%	0.36%	0.37%	0.50%	0.33%	0.38%	0.40%	0.72%	0.57%	0.73%	0.40%	0.44%
60%	0.33%	0.30%	0.32%	0.36%	0.29%	0.34%	0.35%	0.32%	0.33%	0.36%	0.28%	0.34%	0.32%	0.40%	0.37%	0.42%	0.32%	0.42%
70%	0%	0%	0%	0%	0.28%	0.28%	0%	0%	0%	0%	0.28%	0.28%	0.32%	0.39%	0.29%	0.39%	0.32%	0.34%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.23%	0%	0%	0%	0.23%	0.23%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.08%	0.95%	1.79%	6.33%	1.75%	2.56%	2.45%	1.84%	2.23%	5.74%	1.91%	2.69%	2.62%	7.22%	6.50%	7.70%	2.62%	3.50%
AVG100	0.45%	0.40%	0.66%	2.06%	0.66%	0.91%	0.88%	0.68%	0.80%	1.88%	0.72%	0.95%	0.96%	2.40%	2.15%	2.54%	0.96%	1.24%

T438 Query 1

T438 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	2.94%	2.99%	1.56%	4.13%	3.23%	3.97%	4.55%	1.70%	1.44%	4.95%	3.23%	4.55%	4.81%	4.46%	5.21%	4.95%	3.23%	4.31%
20%	1.89%	1.44%	1.08%	4.87%	1.92%	4.62%	3.35%	1.10%	1.15%	5.63%	1.91%	3.35%	4.55%	3.35%	5.96%	5.63%	1.91%	3.29%
30%	1.21%	0.99%	1.02%	3.45%	0.47%	2.81%	3.47%	0.93%	0.92%	1.53%	0.47%	3.47%	2.71%	1.13%	1.27%	1.61%	0.46%	3.69%
40%	0.88%	0.63%	0.51%	1.40%	0.53%	1.41%	0.48%	0.79%	0.50%	1.30%	0.55%	1.75%	1.69%	0.71%	1.31%	1.27%	0.55%	1.77%
50%	0.34%	0.32%	0.33%	0.73%	0.57%	1.04%	0.52%	0.69%	0.31%	0.38%	0.56%	1.28%	1.00%	0.78%	0.37%	0.66%	0.56%	1.48%
60%	0.36%	0.37%	0.30%	0.34%	0.33%	0.98%	0.30%	0.36%	0.26%	0.27%	0.35%	1.00%	0.43%	0.37%	0.27%	0.31%	0.40%	0.97%
70%	0.00%	0.00%	0.00%	0.24%	0.23%	0.24%	0.16%	0.16%	0.15%	0.15%	0.25%	0.19%	0.30%	0.39%	0.15%	0.26%	0.27%	0.20%
80%	0.00%	0.00%	0.00%	0.00%	0.05%	0.05%	0.00%	0.00%	0.00%	0%	0.09%	0.09%	0%	0%	0%	0%	0.09%	0.12%
90%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.00%	0.00%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.00%	0.00%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	2.01%	1.81%	1.22%	4.15%	1.87%	3.80%	3.79%	1.24%	1.17%	4.04%	1.87%	3.79%	4.02%	2.98%	4.14%	4.06%	1.87%	3.76%
AVG100	0.76%	0.67%	0.48%	1.52%	0.73%	1.51%	1.28%	0.57%	0.47%	1.42%	0.74%	1.57%	1.55%	1.12%	1.45%	1.47%	0.75%	1.58%

T438 Query 2

T438 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	4.00%	4.07%	3.97%	6.58%	13.16%	3.79%	4.13%	4.00%	3.97%	6.58%	13.16%	3.79%	6.17%	10.20%	10.20%	6.58%	13.16%	3.79%
20%	2.90%	3.85%	4.23%	8.82%	7.38%	4.57%	3.40%	4.27%	4.23%	8.82%	7.38%	4.57%	7.90%	10.35%	10.47%	8.82%	7.38%	4.57%
30%	4%	5.09%	5.09%	7.49%	4.88%	3.00%	3.84%	5.13%	5.09%	7.49%	4.88%	3.00%	7.78%	5.45%	5.39%	7.41%	4.88%	3.00%
40%	3.64%	3.58%	3.63%	6.84%	5.81%	2.05%	4.47%	3.58%	3.63%	6.84%	5.81%	2.05%	6.08%	5.79%	5.94%	6.79%	5.81%	2.02%
50%	3.49%	3.61%	3.63%	6.55%	3.58%	1.68%	3.47%	3.62%	3.63%	6.55%	3.58%	1.68%	6.05%	6.73%	6.63%	6.50%	3.58%	1.51%
60%	3.74%	3.48%	3.50%	6.32%	1.43%	1.63%	3.79%	3.47%	3.50%	6.32%	1.43%	1.63%	6.31%	4.26%	4.36%	6.28%	1.43%	1.53%
70%	2.37%	1.94%	1.98%	5.15%	1.42%	0.72%	2.40%	1.97%	1.98%	5.15%	1.42%	0.72%	5.16%	3.05%	2.93%	5.10%	1.42%	0.72%
80%	1.48%	1.36%	1.36%	2.50%	1.50%	0.64%	1.64%	1.36%	1.36%	2.50%	1.50%	0.64%	2.65%	2.13%	2.08%	2.45%	1.50%	0.64%
90%	1.40%	0.79%	0.94%	1.41%	1.04%	0.60%	1.34%	0.82%	0.93%	1.41%	1.04%	0.60%	1.56%	0.97%	0.99%	1.36%	1.04%	0.60%
100%	0.83%	0.43%	0.43%	0.81%	0.72%	0.49%	1.05%	0.43%	0.43%	0.80%	0.72%	0.49%	0.69%	0.70%	0.71%	0.75%	0.72%	0.41%
Avg30	3.58%	4.33%	4.43%	7.63%	8.47%	3.79%	3.79%	4.46%	4.43%	7.63%	8.47%	3.79%	7.28%	8.67%	8.68%	7.60%	8.47%	3.79%
AVG100	2.77%	2.82%	2.88%	5.25%	4.09%	1.92%	2.95%	2.86%	2.87%	5.25%	4.09%	1.92%	5.03%	4.96%	4.97%	5.20%	4.09%	1.88%

T438 Query 3

T438 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	8.33%	7.69%	4.81%	8.62%	8.48%	8.93%	5.68%	5.32%	4.81%	8.62%	8.48%	8.93%	9.80%	8.62%	8.62%	8.62%	8.48%	8.93%
20%	4.95%	5.20%	4.79%	11.84%	13.64%	12.16%	4.64%	4.95%	4.79%	11.84%	13.64%	12.16%	8.65%	11.84%	11.69%	11.84%	13.64%	12.16%
30%	6.09%	6.45%	3.47%	1.26%	3.64%	4.97%	3.54%	3.78%	3.44%	1.23%	3.52%	4.95%	2.97%	1.37%	1.22%	1.23%	3.52%	4.95%
40%	1.14%	1.02%	0.80%	0.27%	0.76%	0.90%	1.03%	0.80%	0.78%	0.27%	0.75%	0.90%	0.58%	0.28%	0.27%	0.27%	0.75%	0.90%
50%	1.06%	0.88%	0.23%	0.14%	0.23%	0.38%	1.00%	0.23%	0.22%	0.14%	0.23%	0.38%	0.27%	0.18%	0.14%	0.14%	0.23%	0.38%
60%	0.50%	0.49%	0.21%	0.12%	0.16%	0.11%	0.50%	0.21%	0.21%	0.12%	0.16%	0.11%	0.22%	0.12%	0.12%	0.12%	0.16%	0.11%
70%	0.46%	0.42%	0.12%	0.08%	0.13%	0.11%	0.47%	0.12%	0.12%	0.08%	0.13%	0.11%	0.18%	0.08%	0.08%	0.08%	0.13%	0.11%
80%	0.25%	0.26%	0.08%	0.08%	0.10%	0.11%	0.25%	0.08%	0.08%	0.08%	0.10%	0.11%	0.08%	0.08%	0.08%	0.08%	0.10%	0.11%
90%	0.18%	0.18%	0.08%	0.08%	0.10%	0.09%	0.18%	0.08%	0.08%	0.08%	0.10%	0.09%	0.08%	0.08%	0.08%	0.08%	0.10%	0.09%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	6.46%	6.45%	4.35%	7.24%	8.58%	8.69%	4.62%	4.68%	4.35%	7.23%	8.54%	8.68%	7.14%	7.28%	7.18%	7.23%	8.54%	8.68%
AVG100	2.29%	2.26%	1.46%	2.25%	2.72%	2.78%	1.73%	1.56%	1.45%	2.25%	2.71%	2.77%	2.28%	2.27%	2.23%	2.25%	2.71%	2.77%

T438 Query 4

T438 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.43%	0.36%	0.62%	0.60%	0.49%	0.65%	0.78%	0.66%	0.76%	0.74%	0.51%	0.65%	0.77%	0.65%	0.74%	0.74%	0.51%	0.65%
20%	0.44%	0.40%	0.24%	0.26%	0.16%	0.28%	0.39%	0.29%	0.35%	0.34%	0.21%	0.31%	0.37%	0.31%	0.34%	0.34%	0.21%	0.31%
30%	0.10%	0.13%	0.31%	0.30%	0.11%	0.33%	0.40%	0.32%	0.40%	0.39%	0.18%	0.31%	0.42%	0.31%	0.39%	0.39%	0.19%	0.31%
40%	0.07%	0.07%	0.26%	0.28%	0.06%	0.31%	0.28%	0.28%	0.27%	0.27%	0.07%	0.28%	0.27%	0.28%	0.27%	0.27%	0.15%	0.28%
50%	0.07%	0.09%	0.13%	0.13%	0.06%	0.14%	0.10%	0.09%	0.10%	0.12%	0.04%	0.09%	0.19%	0.10%	0.19%	0.19%	0.05%	0.10%
60%	0.06%	0.07%	0.11%	0.13%	0.06%	0.13%	0.08%	0.05%	0.08%	0.12%	0.05%	0.05%	0.12%	0.07%	0.13%	0.13%	0.05%	0.07%
70%	0.06%	0.06%	0.08%	0.13%	0.06%	0.14%	0.08%	0.04%	0.07%	0.09%	0.05%	0.05%	0.12%	0.06%	0.12%	0.12%	0.05%	0.06%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.04%	0.04%	0%	0%	0%	0%	0.04%	0.04%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.33%	0.30%	0.39%	0.38%	0.25%	0.42%	0.52%	0.42%	0.50%	0.49%	0.30%	0.42%	0.52%	0.42%	0.49%	0.49%	0.30%	0.42%
AVG100	0.12%	0.12%	0.18%	0.18%	0.10%	0.20%	0.21%	0.17%	0.20%	0.21%	0.12%	0.18%	0.23%	0.18%	0.22%	0.22%	0.12%	0.18%

T438 Query 5

T438 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	7.25%	1.31%	4.10%	11.11%	38.46%	11.63%	3.60%	7.94%	3.57%	11.11%	14.29%	17.86%	50%	50%	50%	50%	50%	50%
20%	1.33%	1.31%	2.62%	9.18%	1.31%	12%	2.35%	4.25%	2.38%	12.86%	13.85%	5.11%	5.11%	4.89%	5.29%	5.29%	5.52%	4.89%
30%	0.55%	0.55%	2.12%	9.93%	1.94%	4.17%	1.78%	0.92%	1.79%	3.24%	3.57%	6.28%	5.07%	5.05%	5.07%	5.07%	5.02%	5.05%
40%	0%	0%	2.02%	8.74%	2.09%	0.69%	1.80%	0.64%	1.79%	3.46%	2.95%	1.81%	5.06%	4.97%	5%	5%	4.88%	4.97%
50%	0%	0%	1.88%	7.69%	2.11%	0.85%	1.44%	0.75%	1.44%	2.86%	2.84%	1.44%	5.60%	5.46%	5.56%	5.56%	5.46%	5.48%
60%	0%	0%	1.33%	6.86%	2.18%	1.00%	1.50%	0.76%	1.53%	1.93%	1.92%	1.24%	4.61%	4.52%	4.57%	4.57%	4.52%	4.52%
70%	0%	0%	0.15%	0.38%	1.63%	1.04%	1.03%	0.74%	1.04%	0.77%	0.86%	1.07%	2.03%	1.97%	2.00%	1.99%	1.98%	1.98%
80%	0%	0%	0%	0%	0.05%	0.04%	0.14%	0.58%	0%	0.18%	0.33%	0.38%	0.28%	0.27%	0.27%	0.27%	0.33%	0.27%
90%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0.04%	0.03%	0.03%	0.05%	0.05%	0.05%	0.05%	0.06%	0.05%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	3.04%	1.06%	2.95%	10.07%	13.90%	9.27%	2.57%	4.37%	2.58%	9.07%	10.57%	9.75%	20.06%	19.98%	20.12%	20.12%	20.18%	19.98%
AVG100	0.91%	0.32%	1.42%	5.39%	4.98%	3.14%	1.37%	1.66%	1.37%	3.64%	4.06%	3.52%	7.78%	7.72%	7.78%	7.78%	7.78%	7.72%

T438 Query 6

T438 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.16%	0.16%	0.65%	2.72%	2.81%	4.20%	0.62%	1.48%	0.41%	3.38%	2.86%	4.17%	3.38%	4.17%	3.38%	3.38%	2.83%	4.17%
20%	0%	0%	0.21%	2.94%	2.59%	0.26%	0.20%	0.26%	0.39%	2.85%	3.90%	0.26%	2.84%	0.26%	2.86%	2.85%	3.83%	0.26%
30%	0%	0%	0.15%	0.02%	0.02%	0.05%	0.28%	0.29%	0.28%	2.80%	0.75%	0.39%	2.70%	0.39%	2.83%	2.79%	0.76%	0.39%
40%	0%	0%	0.06%	0.02%	0.02%	0.03%	0.25%	0.27%	0.24%	0.30%	0.24%	0.49%	0.30%	0.49%	0.29%	0.29%	0.22%	0.49%
50%	0%	0%	0%	0.03%	0%	0%	0.21%	0.23%	0.22%	0.02%	0.10%	0.05%	0.02%	0.05%	0.02%	0.02%	0.09%	0.05%
60%	0%	0%	0%	0%	0%	0%	0.14%	0.16%	0.14%	0.02%	0.05%	0.04%	0.02%	0.04%	0.02%	0.02%	0.05%	0.04%
70%	0%	0%	0%	0%	0%	0%	0.08%	0.08%	0%	0.02%	0.03%	0.03%	0.02%	0.03%	0.02%	0.02%	0.03%	0.03%
80%	0%	0%	0%	0%	0%	0%	0.04%	0.04%	0%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%
90%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.05%	0.05%	0.34%	1.89%	1.80%	1.50%	0.36%	0.68%	0.36%	3.01%	2.50%	1.61%	2.97%	1.61%	3.02%	3.01%	2.47%	1.61%
AVG100	0.02%	0.02%	0.11%	0.57%	0.54%	0.45%	0.18%	0.28%	0.18%	0.94%	0.80%	0.55%	0.93%	0.55%	0.95%	0.94%	0.79%	0.55%

T438 Query 7

T438 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	16.13%	20%	17.24%	17.24%	35.71%	14.71%	13.51%	19.23%	17.24%	17.24%	35.71%	14.71%	13.51%	19.23%	17.24%	17.24%	35.71%	14.71%
20%	20.46%	20%	16.36%	16.36%	13.24%	16.36%	17.65%	19.57%	16.36%	16.36%	13.24%	16.36%	17.65%	19.57%	16.36%	16.36%	13.24%	16.36%
30%	11.77%	13.21%	21.54%	21.54%	8.24%	17.28%	17.95%	18.42%	21.54%	21.54%	8.24%	17.28%	17.95%	18.42%	21.54%	21.54%	8.24%	17.28%
40%	8.74%	12.68%	16.51%	16.51%	5.19%	12.68%	13.64%	14.75%	16.51%	16.51%	5.19%	12.68%	13.64%	14.75%	16.51%	16.51%	5.19%	12.68%
50%	8.36%	7.96%	10.41%	10.41%	4.04%	10.13%	12.11%	9.24%	10.41%	10.41%	4.04%	10.13%	12.11%	9.24%	10.41%	10.41%	4.04%	10.13%
60%	7.67%	6.68%	8.33%	8.33%	1.91%	7.61%	11.29%	8.62%	8.33%	8.33%	1.91%	7.61%	11.29%	8.62%	8.33%	8.33%	1.91%	7.61%
70%	5.84%	4.53%	5.73%	5.73%	1.60%	6.25%	7.75%	7.48%	5.73%	5.73%	1.60%	6.25%	7.75%	7.48%	5.73%	5.73%	1.60%	6.25%
80%	2.35%	2.40%	2.22%	2.22%	1.33%	1.02%	2.31%	1.90%	2%	2.20%	1.33%	1.02%	2.31%	1.90%	2.20%	2.20%	1.33%	1.01%
90%	1.42%	0.87%	1.04%	1.04%	0.98%	0.51%	1.22%	0.97%	1%	0.99%	0.96%	0.40%	1.22%	0.97%	0.99%	0.99%	0.96%	0.40%
100%	0.46%	0.46%	0.41%	0.41%	0.05%	0.38%	0.33%	0.46%	0%	0.32%	0.05%	0.32%	0.33%	0.46%	0.32%	0.32%	0.05%	0.32%
Avg30	16.12%	17.74%	18.38%	18.38%	19.06%	16.12%	16.37%	19.07%	18.38%	18.38%	19.06%	16.12%	16.37%	19.07%	18.38%	18.38%	19.06%	16.12%
AVG100	8.32%	8.88%	9.98%	9.98%	7.23%	8.69%	9.77%	10.06%	9.96%	9.96%	7.23%	8.68%	9.77%	10.06%	9.96%	9.96%	7.23%	8.68%

T438 Query 8

T438 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	4.13%	4.17%	7.81%	14.71%	27.78%	13.51%	6.49%	6.94%	7.46%	14.71%	27.78%	13.51%	14.71%	13.51%	13.89%	14.71%	27.78%	13.51%
20%	6.08%	6.82%	7.03%	13.85%	10.98%	15.25%	5.49%	6.67%	6.87%	13.85%	10.98%	15.25%	10.35%	11.69%	15.25%	13.85%	10.98%	15.25%
30%	4.86%	5.17%	5.47%	11.86%	7.37%	12.28%	5.67%	5.30%	5%	11.86%	7.37%	12.28%	11.67%	7.49%	12.28%	11.86%	7.33%	12.28%
40%	4.20%	4.76%	4.57%	2.83%	4.21%	3.60%	3.98%	4.49%	5%	2.82%	4.21%	3.59%	2.54%	2.52%	3.61%	2.62%	4.14%	3.59%
50%	4.41%	3.69%	3.49%	2.22%	0.40%	3.10%	4.20%	3.44%	3%	2.22%	0.40%	3.10%	2.12%	2.07%	3.12%	2.09%	0.40%	3.10%
60%	3.07%	2.83%	2.66%	2.24%	0.47%	2.31%	3.01%	2.64%	3%	2.23%	0.47%	2.31%	2.09%	2.28%	2.32%	2.12%	0.46%	2.31%
70%	2.92%	2.65%	2.37%	1.28%	0.52%	2.02%	2.72%	2.35%	2%	1.28%	0.52%	2.02%	1.84%	1.25%	2.03%	1.25%	0.51%	2.02%
80%	1.88%	1.52%	1.54%	0.22%	0.25%	0.23%	1.84%	1.54%	2%	0.22%	0.25%	0.23%	0.25%	0.22%	0.23%	0.22%	0.25%	0.23%
90%	0%	0%	0%	0%	0.07%	0.06%	0%	0%	0%	0%	0.07%	0.06%	0%	0%	0%	0%	0.07%	0.06%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	5.02%	5.38%	6.77%	13.47%	15.37%	13.68%	5.88%	6.30%	6.58%	13.47%	15.37%	13.68%	12.24%	10.90%	13.81%	13.47%	15.36%	13.68%
AVG100	3.15%	3.16%	3.49%	4.92%	5.20%	5.24%	3.34%	3.34%	3.43%	4.92%	5.20%	5.24%	4.56%	4.10%	5.27%	4.87%	5.19%	5.24%

T438 Query 9

T438 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.99%	2.30%	1.08%	1.10%	8.77%	1.06%	1.03%	1.05%	1.08%	1.10%	7.94%	1.06%	1.04%	1.06%	1.12%	1.10%	7.94%	1.06%
20%	2.80%	2.79%	1.73%	1.74%	0.10%	1.70%	1.73%	1.70%	1.72%	1.72%	0.10%	1.70%	1.73%	1.72%	1.72%	1.72%	0.10%	1.70%
30%	2.16%	2.22%	2.32%	2.33%	0.15%	2.39%	2.33%	2.31%	2.28%	2.29%	0.15%	2.39%	2.41%	2.39%	2.28%	2.29%	0.15%	2.39%
40%	2.47%	2.57%	2.41%	2.65%	0.19%	2.52%	2.18%	2.23%	2.28%	2.61%	0.19%	2.43%	2.53%	2.66%	2.61%	2.61%	0.19%	2.43%
50%	2.39%	2.31%	2.27%	2.73%	0.24%	2.70%	2.28%	2.20%	2.17%	2.58%	0.24%	2.61%	2.61%	2.71%	2.54%	2.58%	0.24%	2.61%
60%	2.52%	2.14%	2.37%	2.44%	0.29%	2.28%	2.26%	2.20%	2.28%	2.34%	0.29%	2.23%	2.49%	2.40%	2.29%	2.34%	0.29%	2.23%
70%	2.32%	1.63%	1.77%	1.88%	0.32%	1.73%	2.09%	1.70%	1.73%	1.82%	0.32%	1.70%	2.14%	1.81%	1.75%	1.82%	0.32%	1.70%
80%	0.15%	0.16%	0.16%	0.15%	0.08%	0.15%	0.15%	0.15%	0%	0.15%	0.08%	0.15%	0.15%	0.16%	0.15%	0.15%	0.08%	0.15%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	2.32%	2.44%	1.71%	1.72%	3.01%	1.71%	1.70%	1.69%	1.69%	1.70%	2.73%	1.71%	1.73%	1.72%	1.70%	1.70%	2.73%	1.71%
AVG100	1.68%	1.61%	1.41%	1.50%	1.01%	1.45%	1.41%	1.35%	1.37%	1.46%	0.93%	1.43%	1.51%	1.49%	1.44%	1.46%	0.93%	1.43%

T438 Query 10

T438 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.40%	0.08%	0.10%	1.33%	0.22%	1.26%	0.15%	0.20%	0.09%	0.72%	0.22%	1.26%	1.27%	0.89%	0.73%	1.47%	0.22%	1.26%
20%	0.13%	0.13%	0.11%	0.81%	0.32%	0.26%	0.11%	0.22%	0.10%	0.68%	0.32%	0.26%	1.78%	0.22%	0.13%	1.23%	0.35%	0.26%
30%	0%	0%	0.12%	0.15%	0.14%	0.13%	0.12%	0.08%	0.08%	0.14%	0.11%	0.26%	1.56%	0.19%	0.12%	1.68%	0.48%	0.37%
40%	0%	0%	0%	0.14%	0.14%	0.16%	0.09%	0.09%	0.09%	0.10%	0.13%	0.11%	1.23%	0.24%	0.09%	1.22%	0.56%	0.47%
50%	0%	0%	0%	0%	0.03%	0.03%	0.12%	0.12%	0.12%	0.12%	0.12%	0.13%	0.23%	0.24%	0.12%	0.23%	0.17%	0.26%
60%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0.03%	0.07%	0.15%	0.15%	0.14%	0.15%	0.15%	0.16%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.16%	0.16%	0.16%	0.16%	0.11%	0.17%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.18%	0.07%	0.11%	0.76%	0.23%	0.55%	0.13%	0.16%	0.09%	0.51%	0.22%	0.59%	1.54%	0.43%	0.33%	1.46%	0.35%	0.63%
AVG100	0.05%	0.02%	0.03%	0.24%	0.08%	0.18%	0.06%	0.07%	0.05%	0.17%	0.09%	0.21%	0.64%	0.21%	0.15%	0.61%	0.20%	0.30%

T438 Query 1 entropy

T438 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	5.00%	1.89%	1.12%	5.62%	5.62%	3.57%	3.70%	1.29%	1.06%	5.62%	5.62%	4.67%	4.43%	4.03%	6.02%	5.62%	6.02%	3.55%
20%	0.42%	1.80%	1.27%	7.09%	7.09%	4.64%	1.90%	1.14%	1.20%	6.92%	6.92%	3.80%	6.92%	3.60%	7.09%	6.92%	7.09%	3.80%
30%	0.24%	0.58%	1.01%	2.82%	0.48%	4.06%	0.48%	0.77%	0.96%	1.68%	0.48%	3.75%	3.44%	1.93%	2.58%	2.58%	0.48%	4.19%
40%	0.27%	0.28%	0.58%	0.72%	0.49%	1.16%	0.49%	0.70%	0.54%	0.70%	0.50%	1.53%	2.27%	0.98%	0.25%	0.78%	0.49%	1.39%
50%	0.31%	0.31%	0.33%	0.25%	0.37%	0.98%	0.43%	0.73%	0.32%	0.18%	0.39%	1.10%	1.16%	0.86%	0.15%	0.19%	0.43%	1.19%
60%	0.36%	0.34%	0.30%	0.24%	0.27%	0.35%	0.23%	0.37%	0.25%	0.16%	0.23%	1.01%	0.47%	0.37%	0.16%	0.17%	0.23%	0.95%
70%	0.00%	0.00%	0.00%	0.24%	0.09%	0.24%	0.15%	0.16%	0.15%	0.15%	0.12%	0.19%	0.31%	0.38%	0.16%	0.18%	0.13%	0.21%
80%	0.00%	0.00%	0.00%	0.00%	0.05%	0.05%	0.00%	0.00%	0.00%	0%	0.04%	0.09%	0%	0%	0%	0%	0.05%	0.09%
90%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.00%	0.00%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.00%	0.00%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.89%	1.42%	1.13%	5.17%	4.40%	4.09%	2.03%	1.07%	1.07%	4.74%	4.34%	4.07%	4.93%	3.19%	5.23%	5.04%	4.53%	3.85%
AVG100	0.66%	0.52%	0.46%	1.70%	1.45%	1.50%	0.74%	0.52%	0.45%	1.54%	1.43%	1.61%	1.90%	1.22%	1.64%	1.64%	1.49%	1.54%

T438 Query 2 entropy

T438 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	3.50%	4.43%	4.39%	8.33%	6.85%	4.59%	4.07%	4.39%	4.39%	8.33%	6.85%	4.59%	7.94%	10.87%	10.87%	8.33%	6.85%	4.39%
20%	1.94%	2.39%	2.80%	8.41%	7.14%	2.72%	2.38%	2.80%	2.80%	8.41%	7.14%	2.72%	7.56%	12%	12%	8.41%	7.14%	2.79%
30%	3%	2.96%	3.02%	6.86%	6.45%	2.14%	2.59%	3.02%	3.02%	6.86%	6.45%	2.14%	6.51%	8.28%	8.28%	6.86%	6.45%	2.18%
40%	3.16%	3.22%	3.23%	5.96%	3.13%	1.82%	2.58%	3.22%	3.23%	5.96%	3.13%	1.82%	6.72%	7.76%	7.76%	5.96%	3.13%	1.73%
50%	3.49%	3.28%	3.30%	6.55%	1.26%	1.01%	2.81%	3.29%	3.30%	6.55%	1.26%	1.01%	6.91%	6.79%	6.77%	6.55%	1.26%	1.04%
60%	3.91%	2.84%	2.89%	5.70%	1.35%	0.77%	3.00%	2.84%	2.89%	5.70%	1.35%	0.77%	7.39%	7.57%	7.61%	5.70%	1.35%	0.73%
70%	2.68%	2.15%	2.16%	5.90%	1.25%	0.75%	2.22%	2.15%	2.16%	5.90%	1.25%	0.75%	6.00%	5.84%	5.93%	5.90%	1.25%	0.71%
80%	1.55%	1.61%	1.61%	3.17%	1.27%	0.64%	1.55%	1.61%	1.61%	3.17%	1.27%	0.64%	3.51%	3.01%	3.03%	3.16%	1.27%	0.65%
90%	1.39%	0.78%	0.79%	2.27%	0.92%	0.61%	0.79%	0.78%	0.79%	2.26%	0.92%	0.61%	2.01%	1.48%	1.48%	2.26%	0.92%	0.62%
100%	0.51%	0.43%	0.43%	0.21%	0.35%	0.62%	0.44%	0.43%	0.43%	0.21%	0.35%	0.62%	1.08%	0.43%	0.43%	0.13%	0.27%	0.59%
Avg30	2.74%	3.26%	3.40%	7.87%	6.81%	3.15%	3.01%	3.40%	3.40%	7.87%	6.81%	3.15%	7.34%	10.38%	10.38%	7.87%	6.81%	3.12%
AVG100	2.49%	2.41%	2.46%	5.34%	3.00%	1.57%	2.24%	2.45%	2.46%	5.34%	3.00%	1.57%	5.56%	6.40%	6.42%	5.33%	2.99%	1.54%

T438 Query 3 entropy

T438 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	5.10%	5.56%	7.14%	8%	10.64%	9.09%	7.81%	7.58%	7.14%	7.81%	10.64%	9.09%	8.07%	8.07%	7.81%	7.81%	10.64%	9.09%
20%	7.69%	7.50%	3.67%	13.04%	10.59%	13.24%	10.59%	4.04%	3.63%	12.86%	10.59%	13.24%	13.04%	12.68%	12.86%	13.24%	10.59%	13.24%
30%	7.49%	7.65%	3.70%	3.70%	2.59%	4.40%	3.50%	4.40%	3.60%	3.50%	2.50%	4.39%	3.07%	3.76%	3.49%	3.49%	2.49%	4.38%
40%	1%	1%	0.29%	0.28%	0.18%	0.39%	0.35%	0.29%	0.29%	0.27%	0.18%	0.39%	0.24%	0.26%	0.27%	0.27%	0.18%	0.39%
50%	0%	0%	0.16%	0%	0.15%	0.14%	0.15%	0.16%	0.16%	0.15%	0.15%	0.14%	0.13%	0.14%	0.14%	0.14%	0.15%	0.14%
60%	0%	0%	0%	0%	0%	0.11%	0.10%	0.11%	0.11%	0.10%	0.09%	0.11%	0.11%	0.10%	0.10%	0.10%	0.09%	0.11%
70%	0%	0%	0%	0%	0%	0.11%	0.10%	0.08%	0.08%	0.07%	0.06%	0.11%	0.07%	0.07%	0.07%	0.07%	0.06%	0.11%
80%	0%	0%	0%	0%	0%	0.08%	0%	0%	0.07%	0.08%	0.04%	0.08%	0.08%	0.08%	0.08%	0.08%	0.04%	0.08%
90%	0%	0%	0%	0%	0%	0.04%	0%	0%	0.08%	0.08%	0.04%	0.04%	0.08%	0.08%	0.08%	0.08%	0.04%	0.04%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	6.76%	6.90%	4.84%	8.23%	7.94%	8.91%	7.30%	5.34%	4.79%	8.06%	7.91%	8.91%	8.06%	8.17%	8.05%	8.18%	7.91%	8.90%
AVG100	2.22%	2.24%	1.53%	2.54%	2.44%	2.76%	2.28%	1.68%	1.52%	2.49%	2.43%	2.76%	2.49%	2.52%	2.49%	2.53%	2.43%	2.76%

T438 Query 4 entropy

T438 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.36%	0.27%	0.48%	0.48%	0.26%	0.50%	0.54%	0.49%	0.50%	0.49%	0.20%	0.48%	0.53%	0.48%	0.48%	0.48%	0.20%	0.48%
20%	0.25%	0.21%	0.14%	0.14%	0.04%	0.17%	0.19%	0.13%	0.17%	0.19%	0.10%	0.15%	0.21%	0.15%	0.19%	0.19%	0.10%	0.15%
30%	0%	0%	0%	0%	0.04%	0.14%	0%	0%	0.22%	0.22%	0.06%	0.16%	0.24%	0.17%	0.24%	0.24%	0.07%	0.17%
40%	0%	0%	0%	0%	0.04%	0.14%	0%	0%	0.12%	0.12%	0.04%	0.06%	0.25%	0.11%	0.25%	0.25%	0.04%	0.11%
50%	0%	0%	0%	0%	0%	0.12%	0%	0%	0.10%	0.11%	0.04%	0.05%	0.12%	0.05%	0.11%	0.11%	0.04%	0.05%
60%	0%	0%	0%	0%	0%	0.12%	0%	0%	0.10%	0.11%	0.04%	0.05%	0.11%	0.05%	0.11%	0.11%	0.04%	0.05%
70%	0%	0%	0%	0%	0%	0.11%	0%	0%	0.04%	0.10%	0.05%	0.05%	0.10%	0.05%	0.10%	0.10%	0.05%	0.05%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.04%	0%	0%	0%	0%	0.03%	0.04%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.22%	0.18%	0.25%	0.25%	0.11%	0.27%	0.32%	0.26%	0.30%	0.30%	0.12%	0.26%	0.33%	0.27%	0.30%	0.30%	0.12%	0.27%
AVG100	0.09%	0.07%	0.11%	0.12%	0.05%	0.13%	0.14%	0.10%	0.12%	0.13%	0.06%	0.10%	0.16%	0.11%	0.15%	0.15%	0.06%	0.11%

T438 Query 5 entropy

T438 Q5	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	5.75%	1.23%	1.38%	8.93%	50.00%	12.82%	1.33%	3.31%	1.32%	11.36%	31.25%	12.82%	50%	50%	50%	50%	50%	50%
20%	1.28%	1.14%	2.01%	9.57%	1.30%	12.33%	1.91%	5.11%	1.86%	13.43%	1.26%	12.33%	5.17%	5.11%	5.09%	12.33%	4.97%	5.11%
30%	0.55%	0.55%	2.30%	11.20%	1.84%	4.83%	1.95%	1.27%	1.97%	3.02%	1.78%	4.83%	5.83%	5.79%	5.81%	5.81%	5.56%	5.76%
40%	0%	0%	1.70%	7.17%	2.05%	0.69%	1.53%	0.64%	1.54%	3.30%	2.10%	0.69%	4.62%	4.49%	4.58%	4.58%	4.46%	4.49%
50%	0%	0%	1.55%	4.69%	2.01%	0.88%	1.47%	0.74%	1.45%	2.32%	2.12%	0.87%	4.73%	4.61%	4.64%	4.63%	4.35%	4.61%
60%	0%	0%	1%	4%	2%	1.04%	1.40%	0.79%	1.38%	1.42%	2.03%	1.02%	3.64%	3.55%	3.60%	3.60%	3.49%	3.55%
70%	0%	0%	0%	0%	2%	1.11%	1.06%	0.76%	1.02%	0.65%	2.20%	1.07%	1.36%	1.37%	1.37%	1.37%	1.32%	1.38%
80%	0%	0%	0%	0%	0%	0.04%	0%	0%	0%	0.13%	1.39%	0.73%	0.23%	0.22%	0.22%	0.22%	0.25%	0.22%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.04%	0.03%	0.03%	0.05%	0.05%	0.05%	0.05%	0.06%	0.05%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	2.52%	0.97%	1.90%	9.90%	17.71%	9.99%	1.73%	3.23%	1.72%	9.27%	11.43%	9.99%	20.34%	20.30%	20.30%	22.71%	20.18%	20.29%
AVG100	0.76%	0.29%	1.03%	4.57%	6.13%	3.37%	1.08%	1.27%	1.06%	3.57%	4.42%	3.44%	7.56%	7.52%	7.54%	8.26%	7.44%	7.52%

T438 Query 6 entropy

T438 Q6	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.16%	0.16%	0.14%	2.91%	2.56%	4.20%	0.47%	1.08%	0.46%	3.52%	2.86%	4.17%	3.29%	4.17%	3.27%	3.27%	2.84%	4.17%
20%	0.00%	0.00%	0.14%	4.41%	3.09%	0.26%	0.19%	0.26%	0.20%	4.55%	3.90%	0.26%	4.66%	0.26%	4.23%	4.69%	3.83%	0.26%
30%	0.00%	0.00%	0.06%	0.02%	0.02%	0.22%	0.11%	0.17%	0.11%	3.48%	0.84%	0.39%	3.43%	0.39%	3.61%	3.47%	0.83%	0.39%
40%	0.00%	0.00%	0.04%	0.02%	0.02%	0.04%	0.07%	0.07%	0.08%	0.48%	0.23%	0.50%	0.48%	0.49%	0.43%	0.47%	0.23%	0.49%
50%	0.00%	0.00%	0.00%	0.03%	0.00%	0%	0.05%	0.05%	0.04%	0.36%	0.11%	0.14%	0.37%	0.14%	0.10%	0.35%	0.11%	0.14%
60%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.03%	0.05%	0.03%	0.08%	0.03%	0.03%	0.11%	0.03%	0.03%	0.07%	0.03%	0.03%
70%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0.07%	0.02%	0.03%	0.10%	0.03%	0.03%	0.05%	0.02%	0.03%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.04%	0.03%	0.03%	0.04%	0.03%	0.03%	0.03%	0.03%	0.03%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.05%	0.05%	0.11%	2.45%	1.89%	1.56%	0.26%	0.50%	0.25%	3.85%	2.53%	1.61%	3.79%	1.61%	3.70%	3.81%	2.50%	1.61%
AVG100	0.02%	0.02%	0.04%	0.74%	0.57%	0.47%	0.10%	0.18%	0.10%	1.26%	0.80%	0.56%	1.25%	0.56%	1.17%	1.24%	0.79%	0.56%

T438 Query 7 entropy

T438 Q7	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	17.86%	25%	20.83%	20.83%	45.46%	17.24%	15.15%	23.81%	20.83%	20.83%	45.46%	17.24%	15.15%	23.81%	20.83%	20.83%	45.46%	17.24%
20%	20.93%	17.31%	20.46%	20.46%	14.52%	14.29%	18.37%	25%	20.46%	20.46%	14.52%	14.29%	18.37%	25%	20.46%	20.46%	14.52%	14.29%
30%	12.61%	13.08%	19.44%	19.44%	8.38%	18.18%	17.28%	12.50%	19.44%	19.44%	8.38%	18.18%	17.28%	12.50%	19.44%	19.44%	8.38%	18.18%
40%	11.04%	10.17%	14.88%	14.88%	5.11%	14.06%	13.74%	6.04%	14.88%	14.88%	5.11%	14.06%	13.74%	6.04%	14.88%	14.88%	5.11%	14.06%
50%	7.99%	4.46%	10.18%	10.18%	3.49%	9.54%	10.80%	5.26%	10.18%	10.18%	3.49%	9.54%	10.80%	5.26%	10.18%	10.18%	3.49%	9.50%
60%	8.36%	2.60%	8.26%	8.26%	1.21%	6.36%	12.28%	2.19%	8.26%	8.26%	1.21%	6.36%	12.28%	2.19%	8.26%	8.26%	1.21%	6.36%
70%	6.57%	2.60%	5.35%	5.35%	1.14%	4.86%	9.09%	1.68%	5.32%	5.32%	1.14%	4.86%	9.09%	1.68%	5.32%	5.32%	1.14%	4.86%
80%	2.77%	1.66%	2.40%	2.40%	1.10%	1.52%	2.60%	0.95%	2%	2.38%	1.10%	1.52%	2.60%	0.95%	2.38%	2.38%	1.10%	1.51%
90%	1.47%	0.93%	1.20%	1.20%	0.99%	0.52%	1.37%	0.86%	1%	1.17%	0.97%	0.40%	1.37%	0.86%	1.17%	1.17%	0.97%	0.40%
100%	0.46%	0.46%	0.41%	0.41%	0.05%	0.38%	0.36%	0.36%	0%	0.33%	0.05%	0.32%	0.36%	0.36%	0.33%	0.33%	0.05%	0.32%
Avg30	17.13%	18.46%	20.24%	20.24%	22.78%	16.57%	16.93%	20.44%	20.24%	20.24%	22.78%	16.57%	16.93%	20.44%	20.24%	20.24%	22.78%	16.57%
AVG100	9.01%	7.83%	10.34%	10.34%	8.15%	8.70%	10.10%	7.87%	10.32%	10.32%	8.14%	8.68%	10.10%	7.87%	10.32%	10.32%	8.14%	8.67%

T438 Query 8 entropy

T438 Q8	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	5.21%	5.50%	9.43%	11.91%	25%	11.91%	8.48%	8.20%	8.93%	11.11%	25%	11.91%	13.89%	10.87%	11.11%	11.11%	25%	11.91%
20%	7.76%	7.83%	8.49%	13.24%	10.47%	13.04%	7.76%	7.90%	8.26%	12.68%	10.47%	13.04%	15.52%	9.68%	12.50%	12.68%	10.47%	13.04%
30%	6.36%	6.42%	5.86%	16.87%	5.19%	16.67%	5.74%	5.67%	6%	16.28%	5.19%	16.67%	12.39%	6.64%	16.09%	16.28%	5.05%	16.67%
40%	5.81%	6.25%	5.81%	6.59%	2.59%	5.84%	5.39%	5.66%	6%	6.52%	2.59%	5.83%	3.19%	2.80%	5.92%	5.31%	2.56%	5.83%
50%	4.76%	4.11%	3.89%	4.50%	0.40%	4.35%	4.37%	3.83%	4%	4.48%	0.40%	4.34%	2.86%	1.39%	4.37%	3.95%	0.40%	4.35%
60%	3.15%	2.50%	2.42%	2.15%	0.42%	2.11%	2.91%	2.41%	2%	2.15%	0.42%	2.11%	2.22%	0.64%	2.12%	2.04%	0.42%	2.11%
70%	2.55%	2.00%	1.83%	1.91%	0.40%	1.93%	2.48%	1.82%	2%	1.91%	0.40%	1.93%	1.78%	0.29%	1.95%	1.84%	0.40%	1.93%
80%	1.92%	0.22%	0.23%	0.22%	0.16%	0.22%	1.75%	0.23%	0%	0.22%	0.16%	0.22%	0.25%	0.22%	0.22%	0.22%	0.16%	0.22%
90%	0%	0%	0%	0%	0.05%	0.06%	0%	0%	0%	0%	0.05%	0.06%	0%	0%	0%	0%	0.05%	0.06%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	6.44%	6.58%	7.93%	14.00%	13.55%	13.87%	7.32%	7.25%	7.66%	13.36%	13.55%	13.87%	13.93%	9.06%	13.23%	13.36%	13.51%	13.87%
AVG100	3.75%	3.48%	3.80%	5.74%	4.47%	5.61%	3.89%	3.57%	3.71%	5.53%	4.47%	5.61%	5.21%	3.25%	5.43%	5.34%	4.45%	5.61%

T438 Query 9 entropy

T438 Q9	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	1.12%	1.30%	1.09%	1.08%	8.62%	1.09%	1.06%	1.10%	1.08%	1.07%	7.81%	1.09%	1.04%	1.09%	1.07%	1.07%	7.81%	1.09%
20%	1.67%	1.86%	1.77%	1.76%	0.10%	1.76%	1.78%	1.77%	1.73%	1.72%	0.10%	1.76%	1.73%	1.76%	1.72%	1.72%	0.10%	1.76%
30%	2.35%	2.37%	2.25%	2.27%	0.15%	2.30%	2.13%	2.29%	2.09%	2.10%	0.15%	2.30%	2.35%	2.30%	2.09%	2.10%	0.15%	2.30%
40%	2.78%	2.60%	2.62%	2.74%	0.19%	2.77%	2.36%	2.64%	2.44%	2.55%	0.19%	2.77%	2.53%	2.77%	2.55%	2.55%	0.19%	2.77%
50%	2.67%	2.62%	2.67%	2.93%	0.24%	2.95%	2.40%	2.70%	2.53%	2.76%	0.24%	2.95%	2.75%	2.95%	2.76%	2.76%	0.24%	2.95%
60%	2.06%	1.54%	2.79%	2.81%	0.29%	2.80%	2.58%	2.81%	2.66%	2.68%	0.28%	2.80%	2.56%	2.80%	2.66%	2.68%	0.28%	2.80%
70%	1.76%	1.20%	1.65%	1.64%	0.32%	1.59%	2.04%	1.66%	1.61%	1.60%	0.32%	1.59%	1.93%	1.59%	1.59%	1.60%	0.32%	1.59%
80%	0.16%	0.16%	0.16%	0.16%	0.09%	0.16%	0.16%	0.16%	0%	0.15%	0.09%	0.16%	0.15%	0.16%	0.16%	0.15%	0.09%	0.16%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.71%	1.84%	1.70%	1.70%	2.96%	1.72%	1.66%	1.72%	1.63%	1.63%	2.69%	1.72%	1.71%	1.72%	1.63%	1.63%	2.69%	1.72%
AVG100	1.46%	1.36%	1.50%	1.54%	1.00%	1.54%	1.45%	1.51%	1.43%	1.46%	0.92%	1.54%	1.50%	1.54%	1.46%	1.46%	0.92%	1.54%

T438 Query 10 entropy

T438 Q10	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	0.21%	0.08%	0.08%	2.03%	0.20%	1.46%	0.14%	0.17%	0.07%	1.26%	0.20%	1.46%	2.28%	0.17%	1.24%	2.46%	0.22%	1.46%
20%	0.13%	0.13%	0.10%	0.13%	0.34%	0.26%	0.11%	0.15%	0.08%	0.13%	0.34%	0.26%	2.39%	0.17%	0.13%	2.28%	0.36%	0.26%
30%	0%	0%	0.12%	0.15%	0.11%	0.13%	0.08%	0.08%	0.08%	0.11%	0.11%	0.25%	2.63%	0.22%	0.11%	2.76%	0.53%	0.40%
40%	0%	0%	0%	0.14%	0.08%	0.16%	0.09%	0.09%	0.09%	0.09%	0.10%	0.12%	1.36%	0.19%	0.10%	0.25%	0.24%	0.50%
50%	0%	0%	0%	0%	0.03%	0.03%	0.12%	0.12%	0.12%	0.12%	0.08%	0.13%	0.21%	0.13%	0.12%	0.18%	0.17%	0.16%
60%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0.03%	0.07%	0.15%	0.15%	0.14%	0.15%	0.16%	0.16%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.16%	0.16%	0.16%	0.16%	0.06%	0.18%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.11%	0.07%	0.10%	0.77%	0.22%	0.62%	0.11%	0.13%	0.08%	0.50%	0.22%	0.66%	2.43%	0.19%	0.49%	2.50%	0.37%	0.71%
AVG100	0.03%	0.02%	0.03%	0.25%	0.08%	0.20%	0.05%	0.06%	0.04%	0.17%	0.09%	0.23%	0.92%	0.12%	0.20%	0.82%	0.17%	0.31%

T438 Query 1 idf

T438 Q1	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	4.95%	1.89%	1.16%	5.44%	3.97%	3.50%	1.68%	1.30%	1.11%	5.16%	3.97%	3.50%	4.46%	3.73%	5.81%	5.44%	3.97%	4.10%
20%	0.42%	1.78%	1.26%	7.03%	1.94%	4.79%	1.10%	1.14%	1.19%	6.43%	1.92%	4.79%	5.46%	4.50%	7.03%	6.87%	1.92%	3.57%
30%	0.24%	0.57%	1.02%	2.73%	0.48%	3.99%	0.93%	0.78%	0.97%	2.40%	0.48%	3.93%	3.86%	1.31%	0.89%	2.50%	0.47%	4.24%
40%	0.27%	0.28%	0.57%	0.76%	0.49%	1.24%	0.49%	0.72%	0.54%	0.83%	0.50%	1.23%	2.21%	0.66%	0.25%	0.81%	0.50%	1.43%
50%	0.31%	0.31%	0.34%	0.25%	0.38%	0.99%	0.43%	0.72%	0.32%	0.32%	0.39%	1.16%	1.19%	0.48%	0.16%	0.22%	0.44%	1.38%
60%	0.36%	0.34%	0.30%	0.24%	0.27%	0.36%	0.23%	0.37%	0.25%	0.18%	0.24%	0.96%	0.51%	0.38%	0.16%	0.19%	0.24%	0.97%
70%	0.00%	0.00%	0.00%	0.24%	0.09%	0.24%	0.15%	0.16%	0.15%	0.15%	0.12%	0.18%	0.32%	0.38%	0.15%	0.18%	0.17%	0.21%
80%	0%	0%	0%	0%	0.05%	0.05%	0%	0%	0%	0%	0.04%	0.09%	0%	0%	0%	0%	0.05%	0.09%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Avg30	1.87%	1.41%	1.15%	5.07%	2.13%	4.09%	1.24%	1.07%	1.09%	4.66%	2.12%	4.07%	4.59%	3.18%	4.58%	4.94%	2.12%	3.97%
AVG100	0.65%	0.52%	0.47%	1.67%	0.77%	1.51%	0.50%	0.52%	0.45%	1.55%	0.77%	1.58%	1.80%	1.14%	1.45%	1.62%	0.78%	1.60%

T438 Query 2 idf

T438 Q2	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	3.57%	4.72%	4.55%	7.69%	6.76%	4.46%	4.07%	4.55%	4.55%	7.69%	6.76%	4.46%	8.20%	10.87%	10.42%	7.69%	6.76%	4.39%
20%	1.92%	2.42%	2.86%	9.18%	7.50%	2.86%	2.52%	2.86%	2.86%	9.18%	7.50%	2.86%	7.20%	11.54%	12.68%	9.18%	7.50%	2.75%
30%	2.79%	2.98%	3.04%	6.25%	6.48%	2.05%	2.60%	3.02%	3.04%	6.25%	6.48%	2.05%	7.04%	8.43%	7.87%	6.25%	6.48%	2.07%
40%	3.16%	3.22%	3.22%	5.46%	3.10%	1.84%	2.60%	3.22%	3.22%	5.46%	3.10%	1.84%	6.74%	7.47%	7.47%	5.46%	3.10%	1.76%
50%	3.45%	3.34%	3.33%	6.07%	1.26%	1.14%	2.86%	3.33%	3.33%	6.07%	1.26%	1.14%	7.59%	6.63%	6.73%	6.07%	1.26%	1.11%
60%	3.90%	3.27%	3.33%	6.11%	1.36%	0.96%	3.04%	3.28%	3.33%	6.11%	1.36%	0.96%	7.61%	7.27%	7.09%	6.11%	1.36%	0.84%
70%	2.66%	2.31%	2.34%	5.48%	1.30%	0.73%	2.24%	2.32%	2.34%	5.48%	1.30%	0.73%	6.11%	6.04%	6.19%	5.48%	1.30%	0.76%
80%	1.56%	1.57%	1.57%	3.33%	1.31%	0.62%	1.53%	1.57%	1.57%	3.33%	1.31%	0.62%	3.41%	2.93%	3.02%	3.32%	1.31%	0.63%
90%	1.37%	0.81%	0.82%	2.20%	0.88%	0.60%	0.80%	0.81%	0.81%	2.20%	0.88%	0.60%	1.98%	1.45%	1.44%	2.19%	0.88%	0.61%
100%	0.56%	0.43%	0.43%	0.22%	0.44%	0.60%	0.44%	0.43%	0.43%	0.22%	0.44%	0.60%	1.21%	0.43%	0.43%	0.14%	0.42%	0.59%
Avg30	2.76%	3.37%	3.48%	7.71%	6.91%	3.12%	3.06%	3.47%	3.48%	7.71%	6.91%	3.12%	7.48%	10.28%	10.32%	7.71%	6.91%	3.07%
AVG100	2.49%	2.51%	2.55%	5.20%	3.04%	1.59%	2.27%	2.54%	2.55%	5.20%	3.04%	1.59%	5.71%	6.31%	6.33%	5.19%	3.04%	1.55%

T438 Query 3 idf

T438 Q3	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	5.44%	5.21%	7.14%	7.94%	10.64%	9.43%	9.43%	7.35%	7.14%	7.81%	10.64%	9.43%	7.94%	7.81%	7.81%	7.81%	10.64%	9.43%
20%	7.44%	8.11%	3.67%	13.43%	10.47%	13.85%	8.11%	4.02%	3.63%	13.24%	10.47%	13.85%	12.68%	13.04%	13.24%	13.24%	10.47%	13.85%
30%	7.29%	7.33%	3.61%	3.68%	2.50%	3.76%	3.76%	4.29%	3.51%	3.49%	2.42%	3.75%	2.95%	3.66%	3.49%	3.48%	2.41%	3.75%
40%	0.73%	0.61%	0.29%	0.29%	0.18%	0.36%	0.38%	0.29%	0.29%	0.28%	0.18%	0.36%	0.24%	0.27%	0.28%	0.28%	0.18%	0.36%
50%	0.54%	0.36%	0.16%	0.15%	0.15%	0.13%	0.15%	0.16%	0.16%	0.15%	0.15%	0.13%	0.13%	0.15%	0.15%	0.15%	0.15%	0.13%
60%	0.27%	0.29%	0.11%	0.11%	0.10%	0.11%	0.10%	0.11%	0.11%	0.11%	0.10%	0.11%	0.11%	0.11%	0.11%	0.11%	0.09%	0.11%
70%	0.21%	0.21%	0.08%	0.07%	0.07%	0.12%	0.11%	0.08%	0.08%	0.07%	0.07%	0.12%	0.08%	0.07%	0.07%	0.07%	0.07%	0.12%
80%	0.15%	0.16%	0.07%	0.08%	0.04%	0.07%	0.07%	0.07%	0.07%	0.08%	0.04%	0.07%	0.08%	0.08%	0.08%	0.08%	0.04%	0.07%
90%	0.15%	0.12%	0.08%	0.08%	0.04%	0.04%	0.08%	0.08%	0.08%	0.08%	0.04%	0.04%	0.08%	0.08%	0.08%	0.08%	0.04%	0.04%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	6.72%	6.88%	4.81%	8.35%	7.87%	9.01%	7.10%	5.22%	4.76%	8.18%	7.84%	9.01%	7.85%	8.17%	8.18%	8.18%	7.84%	9.01%
AVG100	2.22%	2.24%	1.52%	2.58%	2.42%	2.79%	2.22%	1.65%	1.51%	2.53%	2.41%	2.79%	2.43%	2.53%	2.53%	2.53%	2.41%	2.79%

T438 Query 4 idf

T438 Q4	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.37%	0.27%	0.48%	0.48%	0.28%	0.51%	0.56%	0.52%	1%	0.51%	0.23%	0.51%	0.55%	0.51%	0.50%	0.50%	0.23%	0.51%
20%	0.26%	0.23%	0.15%	0.16%	0.06%	0.18%	0.20%	0.14%	0.18%	0.20%	0.11%	0.16%	0.23%	0.17%	0.20%	0.20%	0.11%	0.17%
30%	0.07%	0.07%	0.13%	0.14%	0.04%	0.14%	0.24%	0.18%	0.24%	0.23%	0.07%	0.17%	0.25%	0.17%	0.25%	0.25%	0.08%	0.17%
40%	0.06%	0.05%	0.14%	0.14%	0.04%	0.14%	0.21%	0.08%	0.13%	0.14%	0.04%	0.08%	0.25%	0.15%	0.25%	0.25%	0.04%	0.15%
50%	0.05%	0.05%	0.11%	0.11%	0.05%	0.12%	0.11%	0.05%	0.10%	0.11%	0.04%	0.05%	0.12%	0.05%	0.11%	0.11%	0.04%	0.05%
60%	0.05%	0.05%	0.10%	0.11%	0.05%	0.11%	0.11%	0.05%	0.11%	0.11%	0.04%	0.05%	0.11%	0.05%	0.11%	0.11%	0.04%	0.05%
70%	0.06%	0.06%	0.06%	0.12%	0.06%	0.12%	0.05%	0.04%	0.04%	0.10%	0.05%	0.05%	0.10%	0.05%	0.10%	0.10%	0.05%	0.05%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0.03%	0.04%	0%	0%	0%	0%	0.03%	0.04%
90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.23%	0.19%	0.25%	0.26%	0.13%	0.28%	0.33%	0.28%	0.31%	0.31%	0.13%	0.28%	0.34%	0.28%	0.32%	0.32%	0.14%	0.28%
AVG100	0.09%	0.08%	0.12%	0.13%	0.06%	0.13%	0.15%	0.11%	0.13%	0.14%	0.06%	0.11%	0.16%	0.12%	0.15%	0.15%	0.06%	0.12%

T438 Query 5 idf

T438 Q5	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	5.88%	1.24%	1.38%	8.77%	50%	15.15%	1.33%	4.85%	1.33%	11.36%	31.25%	15.15%	50%	50%	50%	50%	50%	50%
20%	1.28%	1.23%	2.06%	9.78%	1.31%	9.89%	1.95%	4.95%	1.92%	13.43%	1.28%	9.89%	5.06%	5.06%	5.06%	5.06%	4.95%	5.06%
30%	0.55%	0.55%	2.25%	11.02%	1.84%	4.78%	1.91%	1.23%	1.91%	2.99%	1.80%	4.78%	5.76%	5.49%	5.74%	5.74%	5.67%	5.49%
40%	0%	0%	1.73%	7.23%	2.09%	0.68%	1.52%	0.64%	1.52%	3.28%	2.18%	0.69%	4.69%	4.55%	4.62%	4.62%	4.56%	4.55%
50%	0%	0%	1.55%	6.20%	2.00%	0.85%	1.53%	0.74%	1.51%	2.23%	2.22%	0.86%	4.68%	4.56%	4.65%	4.65%	4.51%	4.56%
60%	0%	0%	1.32%	4.05%	2.24%	1.00%	1.25%	0.79%	1.40%	1.39%	2.06%	1.00%	3.77%	3.71%	3.72%	3.72%	3.64%	3.72%
70%	0%	0%	0.20%	0.22%	1.78%	1.06%	1.08%	0.74%	1.07%	0.66%	2.21%	1.04%	1.51%	1.38%	1.51%	1.50%	1.33%	1.37%
80%	0%	0%	0%	0%	0.04%	0.04%	0.12%	0.21%	0.13%	0.13%	1.40%	0.73%	0.21%	0.20%	0.20%	0.19%	0.26%	0.20%
90%	0%	0%	0%	0%	0%	0%	0.03%	0.04%	0%	0.04%	0.03%	0.04%	0.06%	0.06%	0.05%	0.05%	0.06%	0.06%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	2.57%	1.00%	1.90%	9.86%	17.72%	9.94%	1.73%	3.68%	1.72%	9.26%	11.44%	9.94%	20.27%	20.18%	20.26%	20.26%	20.20%	20.18%
AVG100	0.77%	0.30%	1.05%	4.73%	6.13%	3.34%	1.07%	1.42%	1.08%	3.55%	4.44%	3.42%	7.57%	7.50%	7.55%	7.55%	7.50%	7.50%

T438 Query 6 idf

T438 Q6	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.16%	0.16%	0.16%	2.91%	2.86%	4.27%	0.44%	1.27%	0.44%	3.45%	2.99%	4.35%	3.17%	4.35%	3.23%	3.40%	2.94%	4.35%
20%	0%	0%	0.15%	4.29%	3.08%	0.26%	0.20%	0.26%	0.20%	4.50%	4.23%	0.26%	4.64%	0.26%	4.33%	4.43%	4.15%	0.26%
30%	0%	0%	0.15%	0.02%	0.02%	0.24%	0.16%	0.18%	0.15%	3.40%	0.85%	0.39%	3.39%	0.39%	3.34%	3.37%	0.84%	0.39%
40%	0%	0%	0.07%	0.02%	0.02%	0.03%	0.13%	0.13%	0.13%	0.51%	0.26%	0.50%	0.52%	0.50%	0.51%	0.51%	0.25%	0.50%
50%	0%	0%	0%	0.03%	0%	0%	0.12%	0.10%	0.10%	0.39%	0.12%	0.11%	0.39%	0.11%	0.39%	0.39%	0.11%	0.11%
60%	0%	0%	0%	0%	0%	0%	0.07%	0.06%	0.07%	0.22%	0.03%	0.03%	0.26%	0.03%	0.22%	0.22%	0.03%	0.03%
70%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0.03%	0.14%	0.03%	0.03%	0.16%	0.03%	0.14%	0.14%	0.03%	0.03%
80%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0.04%	0.03%	0.03%	0.03%	0.03%	0.04%	0.04%	0.03%	0.03%
90%	0%	0%	0%	0%	0%	0%	0.03%	0.03%	0%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%	0.03%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.05%	0.05%	0.15%	2.40%	1.99%	1.59%	0.27%	0.57%	0.26%	3.78%	2.69%	1.67%	3.73%	1.67%	3.63%	3.74%	2.64%	1.67%
AVG100	0.02%	0.02%	0.05%	0.73%	0.60%	0.48%	0.12%	0.21%	0.12%	1.27%	0.85%	0.57%	1.26%	0.57%	1.22%	1.25%	0.84%	0.57%

T438 Query 7 idf

T438 Q7	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	17.86%	25%	20%	20%	41.67%	17.24%	15.15%	21.74%	20%	20%	41.67%	17.24%	15.15%	21.74%	20%	20%	41.67%	17.24%
20%	20.46%	17.31%	20%	20%	16.98%	14.29%	18.37%	25.71%	20%	20%	16.98%	14.29%	18.37%	25.71%	20%	20%	16.98%	14.29%
30%	12.39%	12.96%	20%	20%	10%	18.18%	17.28%	12.84%	20%	20%	10%	18.18%	17.28%	12.84%	20%	20%	10%	18.18%
40%	11.04%	10%	15.13%	15.13%	7.47%	13.95%	13.74%	6.19%	15%	15.13%	7.47%	13.95%	13.74%	6.19%	15.13%	15.13%	7.47%	13.95%
50%	8.30%	4.64%	10%	10%	3.96%	9.91%	10.80%	5.40%	10%	10%	3.96%	9.91%	10.80%	5.40%	10%	10%	3.96%	9.91%
60%	8.59%	2.65%	8.75%	8.75%	1.30%	6.54%	12.44%	2.25%	9%	8.75%	1.30%	6.54%	12.44%	2.25%	8.75%	8.75%	1.30%	6.54%
70%	6.85%	2.59%	5.46%	5.46%	1.18%	4.85%	9.25%	1.68%	5%	5.42%	1.18%	4.85%	9.25%	1.68%	5.42%	5.42%	1.18%	4.85%
80%	2.80%	1.77%	2.39%	2.39%	1.14%	1.47%	2.60%	0.99%	2%	2.37%	1.14%	1.47%	2.60%	0.99%	2.37%	2.37%	1.14%	1.47%
90%	1.45%	0.92%	1.19%	1.19%	0.96%	0.52%	1.36%	0.86%	1.15%	1.15%	0.93%	0.40%	1.36%	0.86%	1.15%	1.15%	0.93%	0.40%
100%	0.47%	0.46%	0.41%	0.41%	0.05%	0.38%	0.37%	0.36%	0.33%	0.33%	0.05%	0.32%	0.37%	0.36%	0.33%	0.33%	0.05%	0.32%
Avg30	16.90%	18.42%	20.00%	20.00%	22.88%	16.57%	16.93%	20.10%	20.00%	20.00%	22.88%	16.57%	16.93%	20.10%	20.00%	20.00%	22.88%	16.57%
AVG100	9.02%	7.83%	10.33%	10.33%	8.47%	8.73%	10.14%	7.80%	10.32%	10.32%	8.47%	8.72%	10.14%	7.80%	10.32%	10.32%	8.47%	8.72%

T438 Query 8 idf

T438 Q8	ko+ ENTIRE	ko+ BASE	ko+ SUB	ko+ SUPER	ko+ SEP	ko+ SEP_FC	sub+ ENTIRE	sub+ BASE	sub+ SUB	sub+ SUPER	sub+ SEP	sub+ SEP_FC	super+ ENTIRE	super+ BASE	super+ SUB	super+ SUPER	super+ SEP	super+ SEP_FC
10%	5.21%	5.32%	9.43%	11.91%	25%	10.87%	8.33%	8.20%	8.93%	10.87%	25%	10.87%	14.29%	11.36%	11.91%	10.87%	25%	10.87%
20%	7.76%	7.69%	8.41%	13.24%	10.71%	12.86%	7.56%	7.83%	8.18%	12.16%	10.71%	12.86%	15.25%	9.68%	12.50%	12.16%	10.71%	12.86%
30%	6.36%	6.31%	5.81%	16.67%	5.49%	18.18%	5.67%	5.62%	5.74%	16.87%	5.49%	17.95%	12.50%	6.67%	16.09%	16.87%	5.36%	17.95%
40%	5.71%	6.21%	5.77%	5.90%	2.72%	6.90%	5.29%	5.63%	6%	6.74%	2.72%	6.87%	3.19%	2.75%	5.86%	5.49%	2.70%	6.87%
50%	4.75%	4.10%	3.91%	4.29%	0.40%	4.38%	4.36%	3.87%	4%	4.40%	0.40%	4.37%	2.88%	1.45%	4.29%	3.89%	0.40%	4.38%
60%	3.43%	2.43%	2.36%	2.09%	0.43%	2.68%	3.14%	2.35%	2%	2.42%	0.43%	2.68%	2.23%	0.72%	2.09%	2.29%	0.42%	2.68%
70%	2.51%	2.12%	2.05%	2.16%	0.41%	2.09%	2.45%	2.04%	2.05%	2.11%	0.41%	2.09%	1.91%	0.30%	2.16%	2.02%	0.41%	2.09%
80%	1.90%	0.22%	0.25%	0.22%	0.16%	0.22%	1.85%	0.25%	0.25%	0.22%	0.16%	0.22%	0.26%	0.22%	0.22%	0.22%	0.16%	0.22%
90%	0%	0%	0%	0%	0.05%	0.06%	0%	0.00%	0.00%	0.00%	0.05%	0.06%	0.00%	0.00%	0.00%	0.00%	0.05%	0.06%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0.00%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	6.44%	6.44%	7.88%	13.94%	13.73%	13.97%	7.19%	7.22%	7.62%	13.30%	13.73%	13.89%	14.01%	9.24%	13.50%	13.30%	13.69%	13.89%
AVG100	3.76%	3.44%	3.80%	5.65%	4.54%	5.82%	3.87%	3.58%	3.71%	5.58%	4.54%	5.80%	5.25%	3.31%	5.51%	5.38%	4.52%	5.80%

T438 Query 9 idf

T438 Q9	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	1.89%	1.80%	1.08%	1.08%	8.62%	1.08%	1.03%	1.09%	1%	1.07%	7.81%	1.08%	1.05%	1.08%	1.06%	1.07%	7.81%	1.08%
20%	1.66%	1.74%	1.79%	1.78%	0.10%	1.77%	1.73%	1.78%	2%	1.74%	0.10%	1.77%	1.77%	1.77%	1.74%	1.74%	0.10%	1.77%
30%	2.35%	2.39%	2.26%	2.27%	0.15%	2.33%	2.25%	2.29%	2%	2.10%	0.15%	2.33%	2.33%	2.33%	2.09%	2.10%	0.15%	2.33%
40%	2.78%	2.80%	2.61%	2.73%	0.19%	2.76%	2.28%	2.63%	2%	2.55%	0.19%	2.76%	2.53%	2.76%	2.55%	2.55%	0.19%	2.76%
50%	2.65%	2.76%	2.65%	2.92%	0.24%	2.94%	2.47%	2.67%	2.51%	2.75%	0.24%	2.94%	2.71%	2.94%	2.75%	2.75%	0.24%	2.94%
60%	2.05%	1.62%	2.79%	2.81%	0.29%	2.82%	2.42%	2.81%	2.66%	2.68%	0.28%	2.82%	2.60%	2.82%	2.67%	2.68%	0.28%	2.82%
70%	2.04%	1.36%	1.61%	1.61%	0.32%	1.62%	1.93%	1.62%	1.57%	1.57%	0.32%	1.61%	2.07%	1.62%	1.57%	1.57%	0.32%	1.61%
80%	0.16%	0.16%	0.16%	0.16%	0.09%	0.16%	0.16%	0.16%	0.16%	0.15%	0.09%	0.16%	0.15%	0.16%	0.16%	0.15%	0.09%	0.16%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	1.97%	1.98%	1.71%	1.71%	2.96%	1.72%	1.67%	1.72%	1.64%	1.64%	2.69%	1.72%	1.71%	1.72%	1.63%	1.64%	2.69%	1.72%
AVG100	1.56%	1.46%	1.49%	1.53%	1.00%	1.55%	1.43%	1.50%	1.43%	1.46%	0.92%	1.55%	1.52%	1.55%	1.46%	1.46%	0.92%	1.55%

T438 Query 10 idf

T438 Q10	**ko+ ENTIRE**	**ko+ BASE**	**ko+ SUB**	**ko+ SUPER**	**ko+ SEP**	**ko+ SEP_FC**	**sub+ ENTIRE**	**sub+ BASE**	**sub+ SUB**	**sub+ SUPER**	**sub+ SEP**	**sub+ SEP_FC**	**super+ ENTIRE**	**super+ BASE**	**super+ SUB**	**super+ SUPER**	**super+ SEP**	**super+ SEP_FC**
10%	0.25%	0.08%	0.08%	1.98%	0.21%	1.44%	0.14%	0.18%	0.07%	1.23%	0.20%	1.44%	2.24%	0.27%	1.21%	2.43%	0.21%	1.45%
20%	0.13%	0.13%	0.11%	0.19%	0.34%	0.26%	0.11%	0.16%	0.09%	0.18%	0.34%	0.26%	2.33%	0.16%	0.13%	2.20%	0.36%	0.26%
30%	0%	0%	0.12%	0.15%	0.12%	0.13%	0.08%	0.08%	0.08%	0.13%	0.11%	0.25%	2.54%	0.19%	0.11%	2.71%	0.53%	0.40%
40%	0%	0%	0%	0.14%	0.08%	0.16%	0.09%	0.09%	0.09%	0.09%	0.10%	0.12%	1.34%	0.24%	0.10%	0.35%	0.24%	0.50%
50%	0%	0%	0%	0%	0.03%	0.03%	0.12%	0.12%	0.12%	0.12%	0.08%	0.13%	0.21%	0.18%	0.12%	0.21%	0.17%	0.16%
60%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.03%	0.07%	0.15%	0.15%	0.14%	0.15%	0.16%	0.16%
70%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0.16%	0.16%	0.16%	0.16%	0.06%	0.18%
80%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
90%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Avg30	0.13%	0.07%	0.10%	0.77%	0.22%	0.61%	0.11%	0.14%	0.08%	0.51%	0.22%	0.65%	2.37%	0.21%	0.48%	2.45%	0.37%	0.70%
AVG100	0.04%	0.02%	0.03%	0.25%	0.08%	0.20%	0.05%	0.06%	0.04%	0.17%	0.09%	0.23%	0.90%	0.13%	0.20%	0.82%	0.17%	0.31%

Appendix D: 50 topic queries

Topic No	Query		Topic No	Query
401	foreignminorities,Germany		426	lawenforcement,dogs
402	behavioralgenetics		427	UVdamage,eyes
403	osteoporosis		428	decliningbirthrates
404	Ireland,peacetalks		429	Legionnaires'disease
405	cosmicevents		430	killerbeeattacks
406	Parkinson'sdisease		431	robotictechnology
407	poaching,wildlifepreserves		432	profiling,motorists,police
408	tropicalstorms		433	Greek,philosophy,stoicism
409	legal,PanAm,103		434	Estonia,economy
410	Schengenagreement		435	curbingpopulationgrowth
411	salvaging,shipwreck,treasure		436	railwayaccidents
412	airportsecurity		437	deregulation,gas,electric
413	steelproduction		438	tourism,increase
414	Cuba,sugar,exports		439	inventions,scientificdiscoveries
415	drugs,GoldenTriangle		440	childlabor
416	ThreeGorgesProject		441	Lymedisease
417	creativity		442	heroicacts
418	quilts,income		443	U.S.,investment,Africa
419	recycle,automobiletires		444	supercriticalfluids
420	carbonmonoxidepoisoning		445	womenclergy
421	industrialwastedisposal		446	tourists,violence
422	art,stolen,forged		447	Stirlingengine
423	Milosevic,MirjanaMarkovic		448	shiplosses
424	suicides		449	antibioticsineffectiveness
425	counterfeitingmoney		450	KingHussein,peace

Appendix E: Baseline test results

	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%	avg30	avg100
T401	3.11%	5.14%	7.78%	9.78%	12.11%	13.78%	15.92%	17.56%	2.07%	1.40%	**5.34%**	**8.86%**
T402	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T403	90.00%	85.71%	79.41%	73.47%	67.65%	41.99%	38.55%	0.00%	0.00%	0.00%	**85.04%**	**47.68%**
T404	57.14%	5.84%	7.30%	7.45%	7.78%	5.82%	4.99%	3.19%	3.41%	0.00%	**23.43%**	**10.29%**
T405	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T406	89.00%	89.00%	75.00%	80.00%	42.86%	41.18%	27.59%	0.00%	0.00%	0.00%	**84.33%**	**44.46%**
T407	72.73%	68.18%	58.97%	51.72%	42.22%	36.00%	38.69%	30.30%	30.22%	0.00%	**66.63%**	**42.90%**
T408	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T409	89.00%	89.00%	89.00%	80.00%	83.33%	5.56%	3.00%	0.06%	0.07%	0.03%	**89.00%**	**43.91%**
T410	89.00%	75.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**54.67%**	**16.40%**
T411	66.67%	66.67%	71.43%	63.64%	30.00%	11.58%	6.81%	6.97%	0.00%	0.00%	**68.25%**	**32.37%**
T412	77.78%	58.33%	45.65%	40.58%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**60.59%**	**22.23%**
T413	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T414	89.00%	89.00%	89.00%	89.00%	45.46%	50.00%	41.18%	34.78%	37.50%	5.50%	**89.00%**	**57.04%**
T415	66.67%	71.43%	77.78%	31.03%	3.02%	0.96%	0.42%	0.24%	0.00%	0.00%	**71.96%**	**25.16%**
T416	20.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**6.67%**	**2.00%**
T417	75.00%	64.71%	51.52%	57.90%	49.12%	33.67%	19.21%	5.91%	0.00%	0.00%	**63.74%**	**35.70%**
T418	58.82%	64.52%	72.09%	69.49%	49.52%	0.00%	0.00%	0.00%	0.00%	0.00%	**65.14%**	**31.44%**
T419	21.74%	15.39%	7.57%	6.84%	7.00%	4.05%	3.91%	2.59%	1.78%	0.00%	**14.90%**	**7.08%**
T420	89.00%	75.00%	55.56%	54.55%	47.06%	47.62%	47.83%	0.00%	0.00%	0.00%	**73.19%**	**41.66%**
T421	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T422	6.82%	2.99%	0.86%	0.83%	0.81%	0.89%	1.03%	0.95%	0.02%	0.02%	**3.55%**	**1.52%**
T423	89.00%	0.00%	33.33%	0.00%	6.00%	4.88%	0.00%	1.95%	0.00%	2.21%	**40.78%**	**13.74%**
T424	7.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**2.59%**	**0.78%**
T425	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T426	60.00%	25.00%	4.62%	5.15%	1.30%	0.87%	0.60%	0.62%	0.53%	0.54%	**29.87%**	**9.92%**
T427	13.64%	0.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**4.84%**	**1.45%**
T428	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T429	89.00%	89.00%	89.00%	83.33%	85.71%	77.78%	80.00%	0.00%	0.00%	0.00%	**89.00%**	**59.38%**
T430	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T431	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T432	57.14%	21.88%	15.49%	10.77%	5.00%	4.73%	4.02%	0.95%	0.75%	0.00%	**31.50%**	**12.07%**
T433	89.00%	89.00%	23.64%	10.97%	13.21%	3.59%	0.66%	0.54%	0.56%	0.00%	**67.21%**	**23.12%**
T434	89.00%	91.67%	68.75%	74.58%	76.39%	57.90%	57.04%	26.99%	23.52%	2.55%	**83.14%**	**56.84%**
T435	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T436	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T437	89.00%	75.00%	66.67%	35.29%	35.00%	11.94%	7.35%	4.64%	0.47%	0.20%	**76.89%**	**32.56%**

T438	20.83%	18.00%	18.92%	7.47%	5.23%	4.07%	4.36%	3.21%	2.72%	0.38%	**19.25%**	**8.52%**
T439	4.72%	5.61%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**3.44%**	**1.03%**
T440	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T441	75.00%	50.00%	50.00%	52.17%	44.12%	41.46%	0.00%	0.00%	0.00%	0.00%	**58.33%**	**31.28%**
T442	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T443	89.00%	80.00%	43.75%	37.50%	8.09%	9.42%	10.71%	3.15%	0.18%	0.03%	**70.92%**	**28.18%**
T444	89.00%	89.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**59.33%**	**17.80%**
T445	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T446	60.00%	28.00%	18.18%	5.96%	4.43%	1.84%	0.79%	0.77%	0.69%	0.00%	**35.39%**	**12.07%**
T447	89.00%	89.00%	89.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**89.00%**	**26.70%**
T448	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T449	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
T450	75.00%	75.00%	81.82%	40.00%	45.46%	47.37%	34.43%	37.50%	40.30%	0.00%	**77.27%**	**47.69%**
										Avg	**35.28%**	**17.08%**